Dissertation zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
im Fachbereich Geowissenschaften der Freien Universität Berlin

Terrestrial water storage from GRACE gravity data for hydrometeorological applications

Liangjing Zhang
Berlin, Februar 2017

Institut für Meteorologie
Freie Universität Berlin

1. Gutachter:	Prof. Dr. Maik Thomas
2. Gutachter:	Prof. Dr. Andreas Güntner

Tag der Disputation: 16.01.2017

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Dissertation selbstständig und ohne unerlaubte Hilfe angefertigt zu haben. Es wurden keine anderen als die im Text aufgeführten Hilfsmittel und Quellen verwendet. Ein Promotionsverfahren wurden zu keinem früheren Zeitpunkt an einer anderen Hochschule oder bei einem anderen Fachbereich beantragt.

Liangjing Zhang
Berlin, Februar 2017

Abstract

Since launched in 2002, the Gravity Recovery and Climate Experiment (GRACE) has been proven to be a unique way to monitor total water storage (TWS) variations at large spatial scales (>300 km) by measuring Earth gravity changes and provides valuable information for hydrological and hydro-meteorological applications. In this thesis, globally gridded monthly-mean TWS variations are estimated by applying the state-of-the-art post-processing procedure, which has been evaluated first through a closed-loop environment by means of simulated GRACE-type gravity field time-series. In particular, the median scaling factors calculated from an ensemble of multiple global land model simulations do make the re-scaling more robust against particular weaknesses of a single model. The method to estimate gridded fields of measurement, leakage, and re-scaling errors, which can be used for further estimation of the basin-averaged TWS uncertainties, is also introduced. The TWS variations and error estimates are then applied to assess the accuracy of four global numerical model realizations and to identify the advantages and deficiencies of a certain model. Based on four different validation metrics, it is demonstrated that for the 31 largest discharge basins worldwide all model runs agree with the observations to a very limited degree only, together with large spreads among the models themselves. As a common atmospheric forcing data-set is applied to all hydrological models, it is concluded that those discrepancies are not entirely related to uncertainties in meteorologic input, but instead to the model structure and parametrization, and in particular to the representation of individual storage components with different spatial characteristics in each of the models. TWS as monitored by the GRACE mission is sensitive to the different model physics in individual basins and it could offer helpful insight to modellers for the future improvement of large-scale numerical models of the global terrestrial water cycle. In addition, the TWS variations and error estimates are also applied to assess skill scores of three different ensemble sets of decadal hindcasts performed with the coupled climate model MPI-ESM. Moderately positive skill scores of the initialized hindcasts are obtained both with respect to the zero anomaly forecast and the uninitialized projections in particular for lead year 1 in moderate to high latitudes of the Northern Hemisphere. Changes

in the initialization and increased resolution implemented in the different experiments indeed lead to more skillful initialized hindcasts than in the earlier experiments, thereby documenting improvements of the MPI-ESM decadal climate prediction system during the most recent years.

Kurzfassung

Mit dem Start der Satellitenmission GRACE (Gravity Recovery and Climate Experiment) in 2002 ergab sich die einmalige Möglichkeit Variationen in den grossräumigen kontinentalen Wasserspeichern (TWS - total water storage) durch die Beobachtung des zeitvariablen Schwerefeldes der Erde zu erfassen. Hieraus lassen sich eine Vielzahl an Informationen für hydrologische und hydro-meteorologische Anwendungen ableiten. Durch die Anwendung der aktuell gängigsten Prozessierungsmethoden werden in dieser Arbeit monatliche TWS-Variationen auf einem globalen Gitter abgeleitet. Zum Testen und Optimieren der einzelnen Prozessierungsschritte werden simulierte GRACE-Schwerefelddaten aus einem Ensemble verschiedener globaler hydrologischer Modelle verwendet. Es zeigt sich, dass die notwendige Reskalierung der GRACE Schwerefelder durch die Verwendung eines über das Modellensemble gemittelten Skalierungsfaktors robuster gegenüber Eigenarten einzelner Modelle wird. Zudem werden aus den simulierten GRACE-Schwerefeldern Mess-, Leakage- und Reskalierungsfehler für die Abschätzung von TWS-Ungenauigkeiten einzelner Einzugsgebiete bestimmt. Die TWS-Variationen und Fehlerfelder werden anschliessend genutzt um die Genauigkeit von vier globalen numerischen Modellrealisationen sowie deren individuelle Vor- und Nachteile zu untersuchen. Basierend auf vier verschiedenen Validationsmetriken, zeigen die Modellsimulationen für die 31 grössten Einzugsgebiete nur eine sehr begrenzte übereinstimmung untereinender, sowie mit den GRACE Beobachtungen. Da alle hydrologischen Modelle mit den gleichen Atmosphärendaten angetrieben wurde, sind die Diskrepanzen auf verschiedene Modellparametrisierung zurückzuführen. Inbesondere wirkt sich die Modellierung einzelner Speicherkompartimente mit unterschiedlichen räumlichen Charakteristiken eines jeden Modells stark auf das Ergebnis aus. Da TWS-Bebobachtungen von GRACE äusserst sensitiv auf die individuelle Modellphysik in den einzelnen Einzuggebieten reagieren, dient die vorliegende Arbeit zur Validation und Weiterentwicklung grossskaliger numerischer Modelle des globalen Wasserkreislaufs. Ausserdem werden diese TWS-Variationen und Fehlerabschätzungen zur Bestimmung so genannter Skill-Scores dreier verschiedener Ensembles-Sets von dekadischen Hincast-Modellen verwendet, welche mit dem gekoppelten Klimamodell MPI-ESM simuliert

wurden. Leicht positive Skill-Scores erhält man mit der zero-anomaly Vorhersage und den uninitialisierten Projektionen, insbesondere für das erste Jahr in den hohen nördlichen Breiten. Die stetige Verbesserung der Skill-Scores der initialisierten Vorhersage durch änderungen der Initialisierung und Erhöhung der räumlichen und zeitlichen Auflösung dokumentiert eine deutliche Weiterentwicklung des MPI-ESM Models für dekadische Klimavorhersagen.

Contents

List of Figures

List of Tables

List of Acronyms

AET	Actual Evapotranspiration.
AOD1B	Atmosphere and Ocean De-Aliasing Level-1B product.
ASD	Amplitude Spectral Density.
BSWB	Basin-Scale Water Balance.
CE	Center of Mass of Solid Earth system.
CESM	Community Earth System Model.
CF	Center of Figure system.
CM	Center of Mass system.
CMAP	Climate Prediction Center Merged Analysis of Precipitation.
CMIP5	5th Phase of the Coupled Model Intercomparison Project.
CNES	Centre National d'Etudes Spatiales.
CPC	Climate Prediction Centre.
CRPS	Continuous Ranked Probability Score.
CRPSS	Continuous Ranked Probability Skill Score.
CRU	Climate Research Unit at the University of East Anglia.
CSR	Center for Space Research, University of Texas.
DEOS	Delft Institute of Earth Observation and Space Systems.
DLR	National aeronautic and space research centre of Federal Republic of Germany.
DORIS	Doppler Orbit determination and Radio positioning Integrated on Satellite.
ECCO	Estimating the Circulation and Climate of the Ocean model.
ECHAM6	European Center-Hamburg Atmosphere Model version 6.
ECMWF	European Centre for Medium-Range Weather Forecast.
ENSO	El Niño Southern Oscillation.
EOF	Empirical Orthogonal Function.
ESM	Earth System Model.
GCM	General Circulaion Model.
GECCO-2	German part of ECCO synthesis for 2002 - 2007.
GFZ	GeoForschungsZentrum Potsdam.
GHM	Global Hydrology Model.
GIA	Glacial Isostatic Adjustment.

GLDAS	Global Land Data Assimilation System.
GPCC	Global Precipitation Climatology Centre.
GPS	Global Positioning System.
GRACE	Gravity Recovery And Climate Experiment.
GRACE-FO	GRACE Follow-On.
GRAIL	Gravity Recovery and Interior Laboratory.
GRGS	Groupe de Recherches de Géodésie Spatiale.
GSFC	Goddard Space Flight Center.
GSM	Spherical Harmonics of Earth's gravity field.
GSWP	Global Soil Wetness Project.
HAMOCC	Hamburg Ocean Carbon Cycle Model.
IB	Inverted Barometric correction.
ICGEM	International Centre for Global Earth Models.
IERS	International Earth Rotation and Reference System Service.
IPCC	Intergovernmental Panel on Climate Change.
ISDC	Information System and Data Centre.
ISI-MPI	Inter-Sectoral Impact Model Intercomparison Project.
ITSG	Institute of Theoretical Geodesy and Satellite Geodesy.
IUGG	International Union of Geodesy and Geophysics.
JPL	Jet Propulsion Laboratory Pasadena.
JSBACH	Jena Scheme for Biosphere Atmosphere Coupling.
KBR	K-Band Range.
LL-SST	Low-Low Satellite-to-Satellite Tracking.
LR	Low Resolution.
LRI	Laser Ranging Interferometer.
LSDM	Land Surface Discharge Model.
LSM	Land Surface Model.
Miklip	Mittelfristige Klimaprognosen.
MPI-ESM	Max Planck Institute Earth System Model.
MPIOM	Max-Planck-Institute for Meteorology Ocean Model.
MR	Medium Resolution.
MSE	Mean Squared Error.
MSSS	Mean Square Skill Score.
MWI	Microwave Instrument.
NCAR	National Center for Atmospheric Research.
NCEP	National Center for Environmental Prediction.
NOAA	National Oceanic and Atmospheric Administration.
OASIS	Ocean Atmosphere Sea Ice Soil.

OBP	ocean bottom pressure.
OCCAM	Ocean Circulation and Climate Advanced Model.
OMCT	Ocean Model for Circulation and Tides.
ORA-S4	Ocean Reanalysis System 4.
PET	Potential Evapotranspiration.
PREM	Prelimenary Reference Earth Model.
PSD	Power Spectrum Density.
RACMO-2	Regional Atmospheric Climate Model.
RMS	root mean square.
SAR	Synthetic Aparture Radar.
SDS	Science Data System.
SH	Spherical Harmonic.
SLR	Satellite Laser Ranging.
SNR	Signal-to-Noise Ratio.
TWS	Terrestrial Water Storage.
UTLS	Upper Troposphere and Lower Stratosphere Region.
VIC	Variable Infiltration Capacity.
WASMOD-M	Water and snow balance model.
WATCH	Inter-Sectoral Impact Model Intercomparison Project.
WaterMIP	Water Model Intercomparison Project.
WFDEI	WATCH Forcing Data ERA Interim.
WGHM	WaterGAP Global Hydrological Model.

Chapter 1

Introduction

1.1 Motivation

As two thirds of human body is composed of water, human's life is directly dependent on drinking water. Water is also consumed for agriculture, industry and domestic use and benefits humans for centuries. Besides, water circulates through the land just as it does through the human body and is a key component of the climate, which has an indirect impact on human life. Only around 2.5 % of the total amount of water available in the world is fresh water. The terrestrial water storage (TWS) includes all water components on and underneath the Earth's surface, i.e., soil moisture (in subsurface unsaturated zones), surface water (streams of varied size, ponds, artificial reservoirs, canals and wetlands), groundwater (in subsurface saturated zones), snow pack, as well as the water contained in biomass. It affects the partitioning of precipitation into evaporation and runoff, partitioning of available energy of the surface between sensible and latent heat (Kleidon *et al.*, 2014); and it plays important role in the terrestrial water budget. Through feedback to the climate system, TWS influences air temperature, precipitation and evapotranspiration and contains the information of future climate through land surface memory (Dirmeyer *et al.*, 2009; Koster *et al.*, 2004b). For instance, TWS affects the atmospheric circulation by means of surface albedo changes and thermal isolation due to snow cover (Koster *et al.*, 2004a; Meehl *et al.*, 2009). A reliable estimate of snow water is useful to improve the predictability of subseasonal atmospheric conditions (Jeong *et al.*, 2013). Soil moisture exhibits a memory for wet and dry anomalies long after these conditions happened in the atmosphere (Seneviratne *et al.*, 2006) and is one of the important sources to forecast summer precipitation in the midlatitudes of the land area (Koster *et al.*, 2010). Groundwater also has a potentially large impact on the low-frequency climate variability by means of its contributions to soil moisture re-charge

and evapotranspiration (Bierkens & van den Hurk, 2007). As a slowly varying component of the Earth system, water storage can be used to improve seasonal predictions, like air temperature in North America (Koster *et al.*, 2010). However, for the near-term prosperity, water has been overly exploited without a long-term vision for planning and management. Besides, anthropogenic activities like water withdrawals and construction of dams already significantly impact the terrestrial water (Palmer *et al.*, 2008), which poses large challenges to both economy and our ecosystem. In another way, water can sometimes also turn to be an enemy, when there is too much (flood) or too little (drought) of it. It causes deaths and injuries directly and lead to significant economic losses. Such extreme events are becoming more common in recent years as a result of climate change (Fischer & Knutti, 2015). Thus, an observational knowledge of the present-day TWS not only describes the current water mass redistribution in the hydrological cycle but also is essential for streamflow forecast (Milly *et al.*, 2005), as well as its potential impacts on extreme events as droughts and floods (Koster *et al.*, 2004b), thus allowing more efficient water resource management.

1.2 Background and current research

Despite of the importance of water storage, a direct measurement of it remains a challenge. Traditional methods to obtain water storage are field-based or in situ measurements (Dorigo *et al.*, 2011). These measurements are quite accurate, however, as point-wise estimates, they are generally sparse. It is possible to achieve large spatial coverage when satellite techniques as imagery and altimetry are applied, but usually only one component of the TWS (surface water or snow) that is limited to the uppermost surface can be measured.

Except these two measurements, three other techniques to estimate TWS are introduced by Troch *et al.* (2007). One is the basin-scale water balance (BSWB) method (Seneviratne *et al.*, 2004; Hirschi *et al.*, 2006), where the data is derived by combined atmospheric and terrestrial water-balances to estimate TWS variations using water vapor content and moisture flux convergence from atmospheric reanalysis data and river discharge measurements. However, the method is highly dependent on the accuracy of the reanalysis data which often contain systematic errors in particular at inter-annual time scales and longer.

Another way to estimate TWS is from hydrological models. The hydrological models fall into two categories: Land Surface Models (LSMs) and Global Hydrology Models (GHMs). LSMs provide the lower boundary conditions for General Circulation Models (GCMs) of the atmosphere and focus on describing the vertical exchange of heat and water by solving the surface energy and water balance. For instance, the variable infiltration capacity (VIC;

Liang *et al.*, 1994) model is a macroscale, semi-distributed LSM that solves water and energy balances. GHMs, however, focus on solving the water balance equation and simulating catchment outlet streamflow to assess water resources or forecast flood (Gudmundsson *et al.*, 2012). GHMs are generally simple conceptual models, transferring precipitation to evaporation and runoff to represent the processes of water fluxes and storage. There are several well-known GHMs from different research groups, e.g., the water and snow balance model (WASMOD-M; Widén-Nilsson *et al.*, 2007). Large differences between the simulated hydrological variables and water fluxes on land areas are found among the models (Dirmeyer *et al.*, 2006). These are due to uncertainties in terms of model structure and physics, parameter values, and meteorological forcing data used as model inputs (Güntner, 2008). Another essential reason is limited observational data at large spatial scales. Several model inter-comparison activities have been conducted in the past. The Global Soil Wetness Project (GSWP; Dirmeyer *et al.*, 2006) investigates a broad range of LSMs driven by common forcing data. The Water Model Intercomparison Project (Haddeland *et al.*, 2011, WaterMIP), which is a joint project from Water and Global Change (WATCH) and GWSP, focuses on both LSMs and GHMs, while the anthropogenic impacts on water balance are also investigated. The Inter-Sectoral Impact Model Intercomparison Project (ISI-MPI; Schewe *et al.*, 2014) compares the results from a multitude of models to highlight shortcomings and inconsistencies of them. These projects have primarily focused on evapotranspiration or soil moisture content. Gudmundsson *et al.* (2012) also compares nine large-scale hydrological models including both LSMs and GHMs to assess their abilities to detect features of the mean annual runoff cycle. In this thesis, the term hydrological model is used for the two different types of models with no distinction between them. TWS is simulated by hydrological models as the sum of the different TWS components, like soil moisture, snow and surface water. Thus, TWS variations from models are also different by which water components are considered and the way they are simulated.

The third way to observe TWS changes is from the Gravity Recovery and Climate Experiment (GRACE). The water variations and transports are so large on Earth, that they actually influence the Earth's gravity. This gravity change can be observed by satellites from space. Launched in March 2002, GRACE measures the month-to-month changes in the gravitational field of the Earth mainly based on inter-satellite range-rate measurements (Tapley *et al.*, 2004b). After removing short-term variability due to tides in atmosphere (Biancale & Bode, 2006), solid earth (Petit & Luzum, 2010) and oceans (Savcenko & Bosch, 2012), as well as due to non-tidal variability in atmosphere and oceans (Dobslaw *et al.*, 2013) from the observations, the resulting gravity changes represent mass transport phenomena in the Earth system, which are, apart from long-term trends, almost exclusively related to the global

water cycle. Although contaminated by errors which require refined post-processing, GRACE provides valuable information on the large-scale TWS variations.

Shortly after the launch of GRACE satellites, hydrological models were frequently applied to explore the value of GRACE. GRACE shows its ability to detect TWS variations in continental hydrology at spatial scales of several hundred kilometers, and at time scales from ten days to one month. Good agreements in terms of the seasonal dynamics and their continental-scale spatial patterns were found between the TWS from GRACE and hydrological models (Güntner, 2008). However, some large differences, for instance, the differences in amplitude and phase of seasonal water storage variations, were found during the comparison. As the improvement of the GRACE processing techniques and application of new ocean and atmosphere models to remove high-frequency signals, noises are significantly reduced of newly released GRACE solutions than earlier releases. There was increasing confidence with the data and the researches generally moved in reverse to evaluate the hydrological models through GRACE data and the differences between the two were used to detect model deficiencies (Syed *et al.*, 2008a; Zeng *et al.*, 2008). It was also suggested by Güntner (2008) to use GRACE to improve hydrological models which was a major step to accept GRACE as a robust source for hydrological studies. GRACE-based TWS variations were then applied to improve models by tuning the parameters (Niu & Yang, 2006; Werth, 2010; Lo *et al.*, 2010) or assimilating GRACE TWS into the models (Zaitchik *et al.*, 2008; Li *et al.*, 2012).

GRACE TWS variations were also applied in various subjects related to hydrology. For instance, the GRACE estimated TWS became a unique source to quantify variability in the large scale water cycle that were previously unobservable. Significant groundwater depletion caused by water withdrawing was confirmed in specific areas such as northwest and northern India, California Central Valley and northern China (Rodell *et al.*, 2009; Tiwari *et al.*, 2009; Feng *et al.*, 2013). Since GRACE has captured all storage components of water at and underneath the land surface, it was possible for the first time to close the hydro-meteorological water budget accurately (Ramillien *et al.*, 2006; Pan *et al.*, 2012). On global domain, several studies focused on the teleconnections between land hydrology and oceanography (Llovel *et al.*, 2010; Phillips *et al.*, 2012). Combined with other observations, like in situ measurements and rainfall data from remote sensing, GRACE provided unique assistance to monitor the drought at certain areas of the world (Leblanc *et al.*, 2009; Famiglietti, 2014). Besides, GRACE-based TWS has also been used to analyze the elastic deformation of the Earth due to hydrological masses (Kusche & Schrama, 2005; van Dam *et al.*, 2007; Tregoning *et al.*, 2009), thereby providing new information on elastic properties of the Earth in different regions.

1.3 Research objectives and outline

Although TWS from GRACE is supposed to be a highly useful variable to validate hydrological and climate models, there have been only a few studies working in this field. Through almost 14 years of development, there has been continuous improvement of the GRACE dataset and expertise in processing. This length of the time-series together with a recently completed reprocessing of the whole GRACE record (Dahle *et al.*, 2012) motivates us to investigate what can be learned from GRACE for the validation of hydrological model representations of the global terrestrial water cycle and also for skill assessment of decadal prediction system. Although there is a mismatch between the spatial resolution of GRACE data and that of hydrological models, the effective spatial resolution can be extrapolated to finer spatial scales through post-processing. Proper post-processing method is first investigated in a closed-loop environment by means of simulated GRACE-type gravity field time-series. Using this state-of-the-art post-processing method, gridded TWS variations and errors are estimated which can be applied not only for the specific research purpose in this thesis but also by non-geodetic users for their own research purposes. The investigation done in this thesis will also contribute to the specification of the planned GFZ level-3 products for hydrometeorological applications.

The thesis is organized as follows: Chapter 2 introduces the GRACE satellite project, from the satellite design to the principle of satellite observations. The processing method and the processing centers are also introduced. Then the post-processing of the GRACE data-set to estimate TWS is illustrated, including a description of the geodetic background of the gravity determination and the detailed post-processing procedures such as the current methods of filtering, de-striping, rescaling, basin averaging. In Chapter 3, the specific method to estimate gridded TWS for hydrometeorological applications in this work is described in detail, where Chapter 3.3 is published as Zhang *et al.* (2016). The proper post-processing strategy is first investigated through a closed-loop environment by means of simulated GRACE-like gravity field time-series. One main focus is on the gridded scaling factors which are estimated from an ensemble of multiple global land model simulations with different physics and different atmospheric forcing. An iterative method to estimate degree-1 terms from GRACE data alone is also introduced, based on Bergmann-Wolf *et al.* (2014b). The total TWS uncertainties considering measurement error, leakage error and rescaling error are also analyzed. After the introduction of the models of interests, the gridded TWS as introduced in Chapter 3 is applied to validate four hydrological models in Chapter 4, published as Zhang *et al.* (2017). Then the topic is moving to the next concern of this thesis, the validation of decadal forecasts in Chapter 5. Based on the preliminary introduction of the hindcast data and validation

metrics, the application of the gridded TWS to verify the MPI-ESM hindcast experiments is further illustrated in Chapter 5.2, which is published as Zhang *et al.* (2015). Finally, a summary of the above work is given in Chapter 6 together with the outlook of the methods to improve the quality of the TWS from GRACE and its application in future.

Bergmann-Wolf, I., **Zhang, L.** and Dobslaw, H. (2014). Global eustatic sea-level variations for the approximation of geocenter motion from GRACE, J. Geod. Sci., 4, 37 – 48, doi:10.2478/jogs-2014-0006.

Zhang, L., Dobslaw, H., Dahle, C., Sasgen, I., and Thomas, M. (2015). Validation of MPI-ESM decadal hindcast experiments with terrestrial water storage variations as observed by the GRACE satellite mission, Meteorol. Z., doi: 10.1127/metz/2015/0596.

Zhang, L., Dobslaw, H., and Thomas, M. (2016): Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications, Geophys J Int., 206 (1): 368 – 378. doi: 10.1093/gji/ggw153.

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., and Thomas, M. (2017): Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, Hydrol. Earth Syst. Sci., 21, 821-837, doi:10.5194/hess-21-821-2017.

Chapter 2

GRACE data processing

2.1 GRACE satellite mission

The GRACE satellite mission was launched in March 2002 as a joint NASA/DLR (Deutsches Zentrum für Luft - und Raumfahrt) project. The two identical satellites of GRACE are flying at an altitude of 450 – 500 km in a near polar orbit, one following the other with a separate distance of around 220 km. The initial height was chosen as a compromise of small drag effect and high resolution of gravity anomalies. The 89.5° orbit inclination was applied to maximize global coverage of the Earth's gravity field. The design life time was 5 years, and the mission objectives were to accurately observe the global gravity field every 30 days, at a spatial resolution of 400 km (Tapley *et al.*, 2004b). Well beyond the first 5 years of operation, the GRACE mission will hopefully be extended until 2017, although the orbital height is continuously reducing (Tapley *et al.*, 2014). A follow-on mission is already on the schedule and planned to be launched in 2018. The concept is basically the same with evolved versions of GRACE microwave instrument (MWI), GPS receiver, and accelerometer while a more precise laser inter-satellite ranging instrument is also going to be tested.

The concept of direct measuring the Earth's gravity from a satellite pair was first suggested in the 1960s (Wolff, 1969). By the conservation of energy, when atmospheric drag and radiation is neglected, the change in gravitational potential energy could be compensated by the kinetic energy terms, for instance, the velocity differences between the two satellites. It is deduced that the relative velocity is proportional to the geopotential difference at the respective locations. Although some parts of this early derivation was over-simplified, as the ignorance of atmospheric drag and radiation, this two-satellites system reduced the complexity of the data processing and has fewer sources of errors than the normally applied method at that time to measure the long-term integrated effects of gravity from position

measurements. Subsequently, some important algorithms for its implementation were further developed by Colombo (1984). Through the years of development, the whole concept became mature enough to be realized as a Low-Low Satellite-to-Satellite Tracking (LL-SST) mission. GRACE is the first realization of such a mission to map the Earth's gravity field. The accuracy of the mission is high enough that it is also able to detect the temporal gravity field variations down to spatial scales of several hundred kilometers. This LL-SST principle is also applied to the Moon. The NASA mission GRAIL (Gravity Recovery and Interior Laboratory) is the lunar analog of the successful GRACE project and determines a higher-resolution lunar gravity field with improved coverage on both the near side and the far side of the Moon (Zuber *et al.*, 2013; Klinger *et al.*, 2014).

When the GRACE satellites fly over an area located on the Earth's surface with a gravity anomaly, the orbit of the two satellites will be perturbed at slightly different time due to different distance to the source. The leading satellite first senses the anomaly that causes a small orbit perturbation, while the trailing satellite feels the same perturbation afterward with a time delay. This causes range changes between the two satellites, which are measured by a two-way, K/Ka-band microwave-ranging system (KBR). The ranging system provides the information of the changes in the satellites' distance, velocity and acceleration in the form of biased ranges, range-rates and range-accelerations separately. The additional perturbations by non-gravitational forces, such as atmospheric drag and pressure of solar radiation are accounted for using on-board high-frequency accelerometer at the center of mass. Besides that, each satellite is also equipped with GPS receivers for precise time-tagging and orbit determination. For mapping the inertial orientation of the satellites, a pair of star camera assemblies are provided for each satellite (Schmidt *et al.*, 2008). From the information of observed changes in the inter-satellite distance, position and acceleration of each satellite, the Earth's gravity field can be determined.

2.2 Data processing procedure and analysis centers

The data processing of the GRACE science data is regularly handled by the three processing centers within the GRACE Science Data System (SDS): the Geoforschungszentrum in Potsdam, Germany (GFZ), the NASA Jet Propulsion Laboratory (JPL), and the Center for Space Research at the University of Texas, Austin (UTCSR). The SDS is designed to perform all gravity field processing steps to produce monthly mean gravity fields. These centers process the data differently in terms of the applied background models, the integrated orbit period, data weights, and the maximum degree of the estimated gravity harmonics (Wouters *et al.*,

2014). The data products are categorized according to the processing level that has been applied. Level-1 products are reprocessed, time-tagged and normal-pointed instrument data including the K-band ranging, accelerometer, star camera and GPS data of both satellites. Level-2 data are the monthly and mean gravity field derived from calibrated and validated GRACE level-1 data product. This is generally done using the "dynamical orbit approach" through an iterative procedure by analyzing the relationship between orbits and forces on GRACE satellites. For a satellite in a geocentric inertial frame, Newton's equation of motion is expressed as:

$$\ddot{\vec{r}} = \vec{f}(t, \vec{r}, \dot{\vec{r}})/m = \vec{a}_g + \vec{a}_{ng}, \tag{2.1}$$

where m is the mass of the satellites, $\vec{r}$, $\dot{\vec{r}}$ and $\ddot{\vec{r}}$ are the vectors of the satellite's position, velocity and acceleration, and $\vec{f}$ is the force exerting accelerations ($\vec{a}$) on the satellite. The subscript "g" denotes gravitational (conservative) acceleration and "ng" denotes accelerations due to non-gravitational (non-conservative) forces. The non-gravitational accelerations for GRACE include effects mainly from atmospheric drag and radiation pressure of the Sun and the Earth, all of which are measured by the accelerometers of the GRACE satellites. The gravitational acceleration is the sum of the contributions from the static Earth's geopotential, tides, non-tidal atmospheric and oceanic variations, and the perturbations due to the celestial bodies, which are expressed by a-priori best-known geopotential models, also called background gravity models (Dahle *et al.*, 2012). The non-tidal oceanic and atmospheric mass variations are obtained from external models to reduce the aliasing effect, which is referred to as de-aliasing. For instance, non-tidal high-frequency atmospheric and oceanic mass variation models are routinely generated at GFZ as so-called Atmosphere and Ocean De-aliasing Level-1B (AOD1B) products which are added to the static gravity model when GRACE monthly gravity field is estimated. The background models are then used to determine the orbit of both satellites, which are compared with the satellite observations (KBR range, range-rate, range-acceleration and GPS-phase measurements). The non-linear relation between the observations and the targeted parameters is linearized with respect to a set of initial parameters. A set of improvements of the initial parameters are calculated by minimizing the difference between observed and predicted orbits estimated in a least squares way for a selected time span. To guarantee a sufficient coverage, the data is sampled every month. Shorter time intervals lead to a smaller number of Stokes coefficients and a lower spatial resolution (Cazenave & Chen, 2010). For several months, the GRACE satellites have a poor ground coverage because of occasional repeat orbits, for instance, when the satellites were in a so-called 61/4 resonance orbit. For these months, regularized and constrained solutions are also provided by the GRACE data centers. By applying a regularization of the equations

of observations based on Kaula's power law (Section 2.3.4, Kaula, 1966), the solutions are constrained by information from a priori background models.

There has been continuous improvement of the processing techniques and the current version of GRACE Level-2 products from SDS is Release 05. Since major differences are found between the GFZ RL05 time series compared to CSR's and JPL's, the whole RL05 time series has been reprocessed and redistributed as RL05a time series.

Besides the SDS-groups, other research groups also provide GRACE solutions, which include the Goddard Space Flight Center (NASA/GSFC; Rowlands *et al.*, 2005), the Delft Institute of Earth Observation and Space Systems (DEOS; Klees *et al.*, 2008), the Groupe de Recherche de Geodesie Spatiale (GRGS; Lemoine *et al.*, 2007), the Institute of Geodesy at University of Graz (Mayer-Gürr *et al.*, 2014; Klinger *et al.*, 2016), the Institute of Theoretical Geodesy (ITG) at the University of Bonn (Eicker, 2008; Mayer-Gürr, 2006), the Astronomical Institute of the University of Bern, Tongji University (Tongji; Chen *et al.*, 2016) and others. Some of these groups also provide 10-day solutions and are working on daily solutions as well. Some groups also estimate water storage variations directly from the level-1B measurements, so called mass concentration block (mascon) solutions. Masses at points or uniform distributed over a certain region are assumed as mascon basis. The mascon basis functions are then related to the range-rate measurements by explicit partial derivatives. There are, however, two ways to express the mascon basis function. One is represented by an analytic expression (Watkins *et al.*, 2005), while the other uses Stokes coefficient representations (Rowlands *et al.*, 2010). The sets of lumped Stokes coefficients are rescaled to fit the K-band range rate observations. As constraint equations from a priori information are applied during the least squares inversion, which largely improves the signal-to-noise ratio, post-processing as smoothing and de-striping is therefore not as important as for the unconstrained spherical harmonic solutions. The monthly gravity field variations in terms of 4,551 equal-area 3-degree spherical cap mascons (JPL RL05M Mascon solution; Watkins *et al.*, 2015) can be downloaded from Tellus website, while a global set of 41,168 1x1 arc-degree equal-area mascons are also provided by the NASA GSFC global mascon products (Luthcke *et al.*, 2013).

2.3 Post-processing of GRACE gravity field models

GRACE level-02 products are classically provided as sets of spherical harmonic coefficients. As unconstrained harmonic solutions from GRACE are contaminated by the noisy short wavelength components and suffer from poor sensitivity to east-west gradients, a post-processing procedure has to be applied to derive reliable surface mass variations from these coefficients.

First, a mathematical description of the gravity potential is needed to represent the gravity field of the Earth.

2.3.1 The Earth's gravity field

According to Newton's gravitational law, the gravitational force between any two particles with masses M and m attracted to each other can be expressed as:

$$F = \frac{GmM}{L^2}. \tag{2.2}$$

G is the universal gravitational constant: $G = 6.6742 \times 10^{(-11)} m^3 kg^{(-1)} s^{(-2)}$. When one of the masses is significantly larger than the other, the smaller mass can be set to unity and larger one can be denoted as M. The gravitational potential V is the potential energy in the field due to M per unit mass and also the integration over space of the gravity field:

$$V = \frac{GM}{L}. \tag{2.3}$$

Vice versa, the gravity field which is a conservative quantity, can be expressed as the spatial derivative of the gravitational potential V. The potential V satisfies Poisson's equation:

$$\mathbf{\Delta} V = -4\pi\rho G, \tag{2.4}$$

where

$$\mathbf{\Delta} V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \tag{2.5}$$

Since outside the attracting masses the density ρ would be zero, 2.4 becomes:

$$\mathbf{\Delta} V = 0. \tag{2.6}$$

This equation 2.6 is called Laplace's differential equation and $\mathbf{\Delta}$ is called the Laplace operator. It is advantageous to use spherical coordinates for Earth, so the Laplace's equation is re-written using the spherical harmonic functions as :

$$\mathbf{\Delta} V = r^2 \frac{\partial^2 V}{\partial r^2} + 2r\frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial \vartheta^2} + \cot\vartheta \frac{\partial V}{\partial \vartheta} + \frac{1}{\sin^2\vartheta} \frac{^2V}{\partial \lambda^2} = 0. \tag{2.7}$$

By solving the Laplace's equation, the potential of gravitational attraction between a unit mass and the Earth system can be expanded into a series of spherical harmonic functions (Heiskanen & Moritz, 1967):

$$V(r,\vartheta,\lambda) = \frac{GM}{r} + \frac{GM}{r}\sum_{l=0}^{N}(\frac{R}{r})^l \sum_{m=0}^{l}[C_{lm}\tilde{P}_{lm}(\cos\vartheta)\cos m\lambda + S_{lm}\tilde{P}_{lm}(\cos\vartheta)\sin m\lambda], \quad (2.8)$$

where r, λ, ϑ are radius, longitude, and co-latitude of the point outside the Earth. R is the semi-major axis of a reference ellipsoid. The term $(\frac{R}{r})^l$ is used here to fulfill being 1 at the surface and being smaller than 1 in the exterior to solve the boundary problem for representing the potential outside Earth's surface. C_{lm} and S_{lm} are the spherical harmonic coefficients with l, m being degree and order. The Legendre functions $\tilde{P}_{lm}(\cos\vartheta)$ are fully normalized as:

$$\tilde{P}_{lm}(\cos\vartheta) = \sqrt{k(2l+1)\frac{(l-m)!}{(l+m)!}}P_{lm}(\cos\vartheta). \quad (2.9)$$

where $k = 1$ when $m = 0$ and $k = 2$ when $m \neq 0$.

2.3.2 Gravity changes due to surface mass

A gravity model consists of numerical values for the Stokes coefficients C_{lm} and S_{lm} . In Wahr *et al.* (1998), a procedure is outlined for using gravity measurements to determine changes in surface mass. If $\rho(r,\vartheta,\lambda)$ is the density redistribution causing the gravity changes, the time-variable Stokes coefficients can be represented by the density redistribution:

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \frac{3}{4\pi R\rho_{ave}(2l+1)} \int \Delta\rho(r,\vartheta,\lambda) \times \tilde{P}_{lm}(\cos\vartheta)(\frac{r}{R})^{l+2} \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \sin\vartheta d\vartheta d\lambda dr. \quad (2.10)$$

where ρ_{ave} is the average density of the Earth and the symbol Δ here is to indicate that we are dealing with anomalies with respect to the long-term average. The mass redistribution can be assumed to happen only in the Earth's surface layer and ρ can therefore be seen as the changes only in this thin layer with a thickness of around 10-15km. Thus, $(\frac{r}{R})^{l+2}$ in Eq. 2.10 equals 1 and the density changes can be interpreted as changes in surface density $\sigma(\lambda,\vartheta)$.

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \frac{3}{4\pi R\rho_{ave}(2l+1)} \int \Delta\sigma(\vartheta,\lambda) \times \tilde{P}_{lm}(\cos\vartheta) \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \sin\vartheta d\vartheta d\lambda. \quad (2.11)$$

The changes of surface mass can cause the direct gravitational changes, while the surface mass variation in loads will cause deformation to the solid Earth which indirectly induces

also gravitational changes. These two influences can be described by the term $1+k_l$ (Farrell, 1972). Consequently, Eq. 2.11 is expanded as follows:

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \frac{3(1+k_l)}{4\pi R\rho_{ave}(2l+1)} \int \Delta\sigma(\vartheta,\lambda) \times \tilde{P}_{lm}(\cos\vartheta) \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \sin\vartheta d\vartheta d\lambda. \quad (2.12)$$

In another way, surface mass changes can be calculated from the changes in the potential coefficients as:

$$\begin{aligned} \Delta\sigma(\vartheta,\lambda) = \frac{R\rho_{ave}}{3} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \frac{2l+1}{1+k_l} \\ \times (\Delta C_{lm}\tilde{P}_{lm}(\cos\vartheta)\cos m\lambda + \Delta S_{lm}\tilde{P}_{lm}(\cos\vartheta)\sin m\lambda). \end{aligned} \quad (2.13)$$

For simplicity, $\Delta\sigma$ could also be expanded from surface density coefficients as:

$$\Delta\sigma(\vartheta,\lambda) = R\rho_w \sum_{l=0}^{\infty} \sum_{m=0}^{l} (\Delta\hat{C}_{lm}\tilde{P}_{lm}(\cos\vartheta)\cos m\lambda + \Delta\hat{S}_{lm}\tilde{P}_{lm}(\cos\vartheta)\sin m\lambda). \quad (2.14)$$

The ratio $\Delta\sigma/\rho_w$ yields the variation in equivalent water height (EWH), which will be used frequently during the analysis of the monthly solutions in the following of the thesis. The relation between the surface density coefficients and the gravitational potential ones can then be expressed as:

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \frac{3\rho_w}{\rho_{ave}} \frac{1+k_l}{2l+1} \left\{ \begin{array}{c} \Delta\hat{C}_{lm} \\ \Delta\hat{S}_{lm} \end{array} \right\}. \quad (2.15)$$

Eq. 2.13 holds true under the assumption that mass redistribution takes place at the sphere of radius R. As the accuracy of GRACE data processing increases, the inaccuracy of this assumption could not be neglected. Instead, it can be assumed that mass transport takes place at the surface of the reference ellipsoid which is closer to the reality. It has been demonstrated by Ditmar (2015) that the ellipsoidal approximation reduces the inaccuracies from 10 % to 1 %. To simplify the computation, the locally spherical approximation could be used. Then the surface mass changes can be expressed as Ditmar (2015):

$$\begin{aligned} \Delta\sigma(\vartheta,\lambda) = \frac{R\rho_{ave}}{3} \sum_{l=0}^{\infty} \sum_{m=0}^{l} (\frac{a}{r(\vartheta,\lambda)})^{l+2} \frac{2l+1}{1+k_l} \\ \times (\Delta C_{lm}\tilde{P}_{lm}(\cos\vartheta)\cos m\lambda + \Delta S_{lm}\tilde{P}_{lm}(\cos\vartheta)\sin m\lambda), \end{aligned} \quad (2.16)$$

where $r(\vartheta,\lambda)$ is radius of the point on the Earth's surface. In practice, the gravity field cannot be estimated with unlimited spatial resolution. Therefore, a certain truncation degree needs

to be set as in Eq. 2.13. The maximum degree $lmax$ relates to the spatial resolution at the Earth's surface as

$$\lambda min \approx 40000km/(lmax + 0.5), \tag{2.17}$$

where λmin is the minimum wavelength of the gravity field. A reasonable maximum degree depends on the precision of the available data, the measurement altitude and also the specific purpose of gravity field modelling.

2.3.3 GRACE Stokes coefficients

The GRACE level-2 data are provided as sets of fully normalized spherical harmonic coefficients: C_{lm}, S_{lm}, also called Stokes coefficients. In this thesis, two current releases of gravity variations provided by GFZ (RL05 and RL05a) are applied, which are published up to degree and order 90 (Dahle *et al.*, 2012). For our thesis, we use the data covering the time span from 01/2003 to 12/2012 where one month is missing due to a 4-day repeat orbit (06/2003) and four are missing due to the battery status of the satellites and switch-offs of the K-band device (01/2011, 06/2011, 05/2012, 10/2012). Atmospheric and oceanic non-tidal mass redistribution and tidal effects are removed during the GRACE data processing using proper background models (Dahle *et al.*, 2012). Stokes coefficients anomalies are obtained by reducing a multi-year average.

As the term ΔC_{00} is proportional to the total mass of the Earth system, it is regarded as a constant due to mass conservation. Degree 1 coefficients are proportional to the position of the center of mass of the Earth relative to the center of the figure (Wahr *et al.*, 1998), which is commonly denoted as geocenter motion (Petit & Luzum, 2010). Geocenter effects are not measured by GRACE due to the fact that the center of mass is also the center of the orbit of the two GRACE satellites. Hence, this data needs to be taken from external sources. Geocenter motion is typically derived from satellite laser ranging (SLR) observations to geodetic satellites. Observations of GPS permanent stations are also used to derive geocenter variations (Blewitt *et al.*, 2001; Fritsche *et al.*, 2010). Besides, joint inversions of several different observations as, satellite altimetry over ocean, in situ ocean bottom pressure observations, GPS permanent station observations and GRACE time-variable gravity fields have also shown promising results (Davis *et al.*, 2004; Rietbroek *et al.*, 2012). For this thesis, the degree 1 terms are estimated from GRACE products alone following the method proposed by Swenson *et al.* (2008). The detailed algorithms are given in Section 4.2.

The degree 2 coefficients are related to the Earth's inertia tensor. The zonal C_{20} coefficient, which is related to the flattening of the Earth, is not well determined by GRACE, because

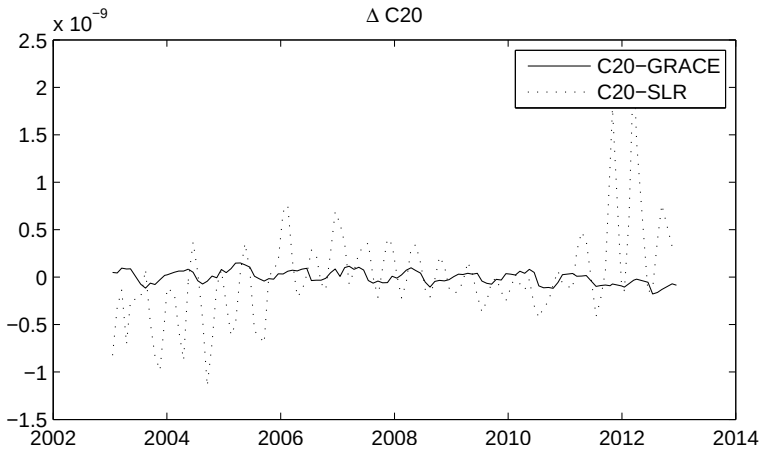

Figure 2.1: The comparison of the C_{20} anomalies from the GRACE SH coefficients and the ones estimated from SLR.

of the orbital geometry of GRACE and the short distance between the two satellites (Tapley *et al.*, 2004a). Therefore, it is typically replaced by external sources, where usually estimates from Satellite Laser Ranging (SLR) are applied. Fig. 2.1 shows the comparison of the C_{20} anomalies with the mean removed from GRACE SH coefficients and the ones estimated from SLR (Cheng *et al.*, 2011). Besides the effect of mass variation due to water storage change within a thin layer at the Earth's surface, GRACE can also detect the mass changes in the Earth's interior. The processes as mantle convection and plate subduction lasts over a very long time scale so that they can be assumed as static over the GRACE period. The effect of glacial-isostatic adjustment (GIA), which is due to Earth's ongoing viscoelastic response to the retreat of the major ice sheets, can cause long-term gravity changes on time scales from 1,000 to 100,000 years. To study hydrological loads on land, the GRACE data are therefore corrected for GIA from a model, given by, e.g., Paulson *et al.* (2007), which is provided in spherical harmonic coefficients ΔC_{lm}^{GIA} ΔS_{lm}^{GIA}. Furthermore, quite large earthquakes, such as the Sumatra-Andaman Earthquake on 24/12/2004, Maule Earthquake (Chile) on 27/2/2010 and Tohoku Earthquake (Japan) on 11/3/2011 cause displacements of the Earth's lithosphere to generate gravity changes that can be detected by GRACE. However, during the GRACE period, most strong earthquakes occurred over the oceans which are not the focus of this thesis, so no correction has been applied.

2.3.4 GRACE uncertainty characteristics

The errors of GRACE potential coefficients generally fall into these categories: 1) Measurement errors of the instruments (caused by system-noise error in KBR range-rate observations,

accelerometer error, and orbit errors), 2) spatial and temporal sampling errors and 3) inaccuracies in the background models, which are investigated in more detail in Schmidt *et al.* (2007). As shown in Eq. 2.8, the external gravitational potential decays inversely with distance from its source, which also depends on the spatial wavelength. Thus the gravity field is dominated by large scale variations while shorter scale variations are damped with increasing distance from the sources. Measurement errors are amplified in the same way and cause the Stokes coefficients uncertainties to become larger with increasing spherical harmonic degree and order. The near-polar orbit with 89.5° inclination leads to an orientation of the ground tracks in North-South direction. As a result, the observations are highly sensitive in this direction, but very limited along the East-West direction. The uncertainties in the background models will cause alias errors, which is further manifested as north-south oriented strips in the spatial gravity field (Fig. 2.2). This is even worse when orbits are repeated in a month or longer time intervals which cause a shortage of data. Fig. 2.3 shows the latitude weighted root mean square (wRMS) of the gridded TWS variations, and we can see the peak values which are caused by the sparse ground track coverage due to occasional repeat orbits.

Signal spectra represents the amplitude of the Earth's gravity field signal. The degree amplitude spectrum expressed as:

$$\Delta\sigma_l = R\sqrt{\sum_{m=0}^{l}(\Delta C_{lm}^2 + \Delta S_{lm}^2)}, \tag{2.18}$$

where R is the spherical earth radius, shows the expected geoid height anomaly amplitude. It includes contributions from all orders at each degree to the total variance of the geoid height anomaly. It is demonstrated that the degree amplitudes meet approximately the so-called Kaula rule (Kaula, 1966):

$$\frac{\sigma_l}{\sqrt{2l+1}} \approx \frac{10^{(-5)}}{l^2}. \tag{2.19}$$

Meantime, the analogous quantity error spectra can also be derived from Eq. 2.18 with the estimated uncertainties in the potential coefficients. Fig. 2.4 shows the amplitude spectrum of both the GRACE signal and errors for the month 04/2003. It is seen that the error of the GRACE Stokes coefficients increases with the degree, corresponding to increased spatial resolution. The formal errors are believed to be too optimistic, particularly for low degrees, which could be caused by the underestimation of the influence of inaccurate background models (Schmidt *et al.*, 2007). To account for this shortcoming, formal errors are calibrated according to Schmidt *et al.* (2008) by applying a degree-dependent scaling factor in order to match certain characteristics of the fields. The calibrated errors provided by GFZ which are the diagonal elements of the covariance matrix are shown as well.

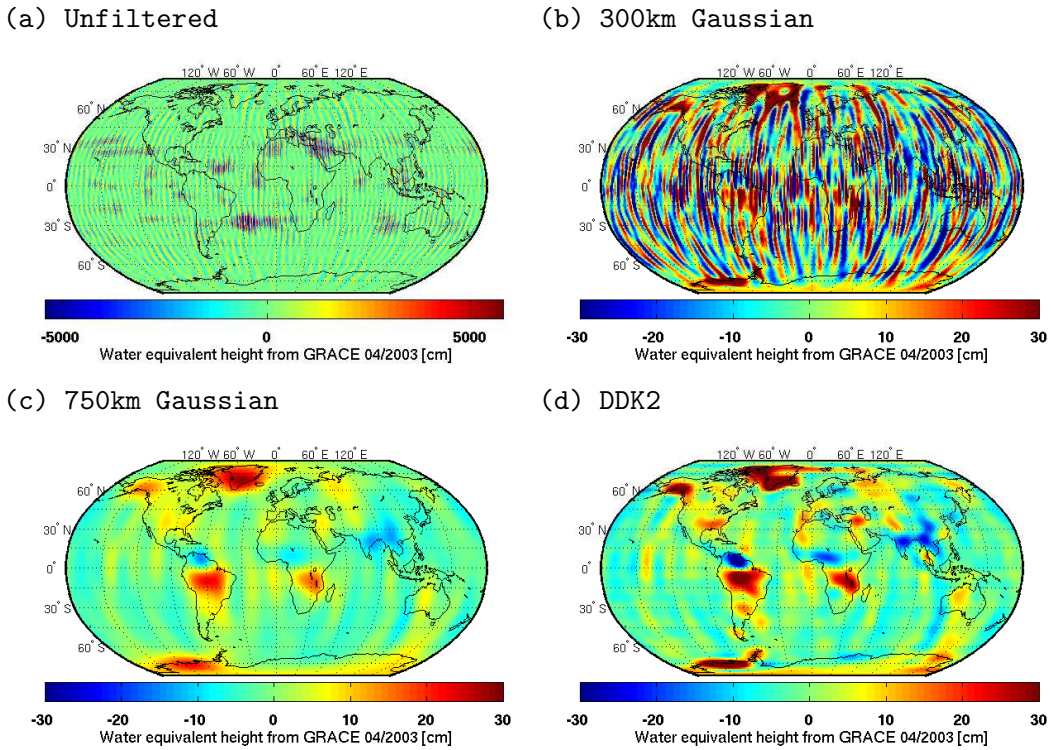

Figure 2.2: Gridded TWS estimated from GRACE GFZ RL05a gravity field when no filter is applied (a), and also the ones filtered with 300km Gaussian (b), 750km Gaussian (c), and DDK2 (d).

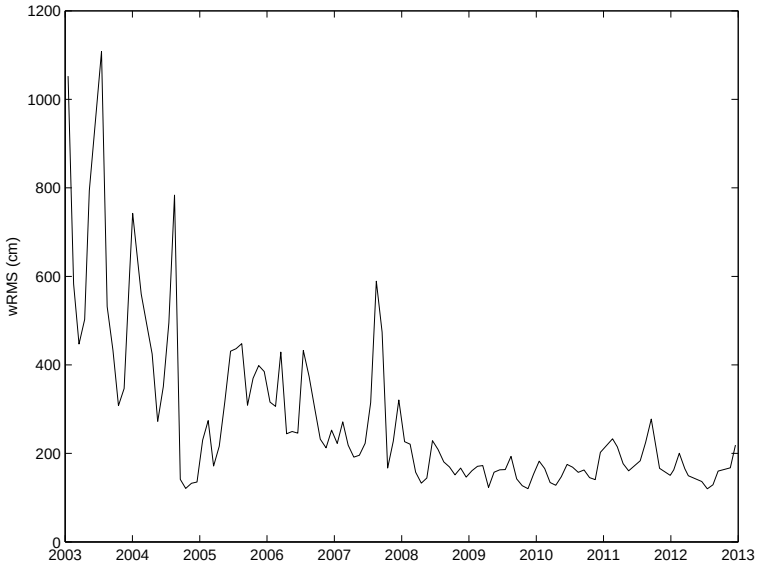

Figure 2.3: Global wRMS of unfiltered gridded TWS variations from GFZ RL05a.

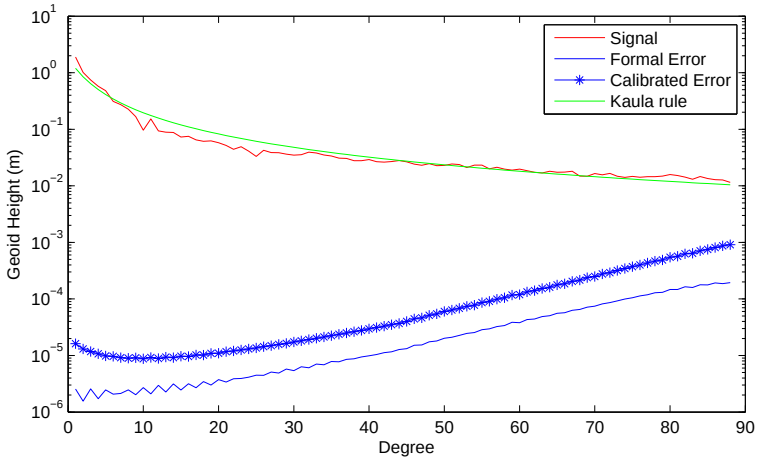

Figure 2.4: The GRACE signal and error degree amplitude for GFZ RL05a gravity field solutions from 04/2003.

2.3.5 Filtering and destriping

As shown above, coefficients of high degree contain large noise and correlated errors. Therefore, it is necessary to filter the Stokes coefficients to balance the trade-off between the gravity field's spatial resolution and its accuracy. The filter coefficients can be applied as:

$$\Delta\sigma(\vartheta,\lambda) = \frac{R\rho_{ave}}{3}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\frac{1+k_l}{2l+1} \times (F_{lm}^{C}\Delta C_{nm}P_{lm}(\cos\vartheta)\cos m\lambda + F_{lm}^{S}\Delta S_{nm}P_{lm}(\cos\vartheta)\sin m\lambda). \tag{2.20}$$

A normally applied method is the isotropic Gaussian filter (Jekeli, 1981), which changes gradually from the highest weight at its center, to half the value at a certain distance on the surface of the Earth which is also denoted as the smoothing radius, then to zero at large distance from the center. The analytical function that approximates the Gaussian filter can be expressed as:

$$W_l = \exp(-\frac{(lr/R)^2}{r\ln 2}), \tag{2.21}$$

while the recursion formulas can be found in Wahr *et al.* (1998). However, the Gaussian kernel is independent of orientation while the noises in the GRACE data have a so-described "striping" pattern. Therefore, a very large filter radius is often needed which causes a large damping and leakage effect.

Alternatively, non-isotropic filtering methods were developed by Swenson & Wahr (2002), Han *et al.* (2005) and Chen *et al.* (2007). Swenson & Wahr (2006) pointed out that the stripes in the surface mass variability estimates are connected with correlations in the spectral domain

for even and odd degree, separately. A quadratic polynomial is thus fitted in a moving window to the Stokes coefficients of like parity and and subtracted from the original Stokes coefficients (appendix A.1). Although the stripes are significantly reduced, the filter performs poorly near the equator. A Gaussian filter therefore needs to be applied afterward. Kusche (2007) combines decorrelation and smoothing with a new filter method: the DDK filter. It imitates the regularization of the normal equations during the GRACE data processing, using a-priori knowledge of the signal and error information (appendix A.2). Fig. 2.2 b, c, and d shows the filtered TWS variations by Gaussian filter for a radius of 300 km (indicated as 300km Gaussian filter), 750km Gaussian and DDK2 for the month 04/2003.

2.3.6 Basin function

Through the data post-processing, the gridded TWS variations are estimated (Fig. 2.2). It should be kept in mind, however, the observations are not point-measurements but rather a spatial average over a certain region due to the indirect observing principle of GRACE. Eq. 2.17 shows the relation between the spatial resolution and the the maximum degree used for spherical harmonic expansion. One can average the gridded TWS over areas of interests:

$$\Delta\sigma_{region} = \frac{1}{\Omega_{region}} \int \Delta\sigma(\vartheta,\lambda) B(\vartheta,\lambda) d\Omega, \tag{2.22}$$

where $d\Omega = \sin\vartheta d\vartheta d\lambda$ is an element of solid angle, and $B(\vartheta,\lambda)$ describes the shape of the target area, with a simplest form given by:

$$B(\vartheta,\lambda) = \begin{cases} 0 & if \quad (\vartheta,\lambda) \in \Omega - R \\ 1 & if \quad (\vartheta,\lambda) \in R \end{cases}, \tag{2.23}$$

where Ω denotes the entire Earth's surface and R the interested region. If the analysis is performed in spectral domain, the basin function can be expressed as:

$$B(\vartheta,\lambda) = \frac{1}{4\pi} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \tilde{P}_{lm}(\cos\vartheta) \cdot (B^C_{lm} \Delta \cos m\lambda + B^S_{lm} \sin m\lambda). \tag{2.24}$$

The basin-averaged water storage over an arbitrary region can be calculated directly from the spherical harmonics of surface density, basin function and the filter:

$$\begin{aligned} \Delta\sigma(\vartheta,\lambda) = \frac{R\rho_{ave}}{3} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \frac{2l+1}{1+k_l} \\ \times (F^C_{lm} B^C_{lm} \Delta C_{nm} \tilde{P}_{lm}(\cos\vartheta) \cos m\lambda + F^S_{lm} B^S_{lm} \Delta S_{nm} \tilde{P}_{lm}(\cos\vartheta) \sin m\lambda) \end{aligned} \tag{2.25}$$

Due to the truncation of the series to degree $lmax$, the basin function does not represent the shape of the target area accurately and ringing appears in the reconstructed basin function paralleling inside and outside of the basin boundary, which is called Gibbs phenomenon (Swenson & Wahr, 2002). Besides, it is seen from Eq. 2.25 that the surface density coefficients can be multiplied with the filter coefficients and the coefficients of the basin function in any order.

2.3.7 Rescaling

The filtering required to reduce GRACE errors will lead to attenuation of the signal amplitude. Besides, water in the surrounding areas will leak into the targeted area, which cause a leakage effect. Because of the bias and leakage problems, there will be under- or overestimation of water storage change, depending on the mass distribution. The leakage effect depends on filter size, basin area and the amplitude of mass variations inside and outside the target area (Klees *et al.*, 2007). The larger the filter, the more signal is leaked outside and the more the surrounding region is sampled. Furthermore, the attenuation effect increases with decreasing basin size. Fig. 2.5a shows the original basin function, Fig. 2.5b the reconstructed basin function using the spherical harmonics truncated to degree $lmax = 90$ and Fig. 2.5c and d the result after filtering the circular function with a Gaussian filter and DDK2 separately. The truncation of the basin function already arises the leakage effect. After filtering, the leakage effect is further amplified. Rescaling can be applied to restore the signal, at basin scale or on the grid. Three methods to remedy the signal alteration are introduced by Long *et al.* (2015). (1) The first method is called scaling-factor approach. A scaling factor is obtained from the least squares fit between the filtered and unfiltered TWS changes from a hydrology model and applied on the filtered TWS changes. (2) The additive correction approach calculates the bias and leakage using output from a hydrological model and obtains the restored TWS as $S_0 = \hat{S}_0 + S_B - S_L$ (Klees *et al.*, 2007; Longuevergne *et al.*, 2010). $\hat{S}_0$ is the filtered GRACE TWS at a catchment:

$$\hat{S}_0 = \frac{1}{R_0} \int_\Omega \hat{S} \cdot B d\Omega. \tag{2.26}$$

S_B and S_L are the bias and leakage respectively which can be estimated from the output S_0 and S_{leak} from a hydrological model. They are expressed in the following equations assuming that the filter is applied on the basin function:

$$S_B = \frac{1}{R_0} \int_R S_0 (B - \hat{B}) d\Omega, \tag{2.27}$$

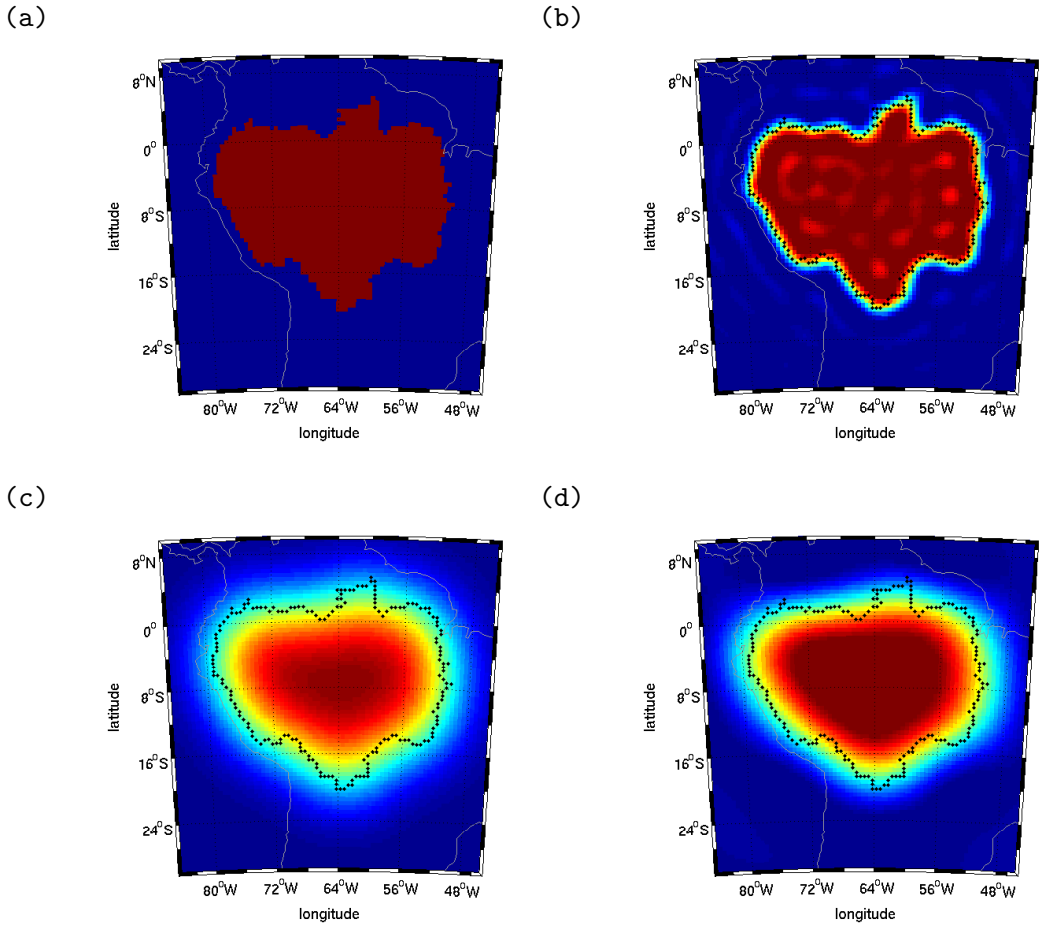

Figure 2.5: The exact basin function for Amazon basin (a), the reconstructed basin function using the spherical harmonics truncated to degree $lmax = 90$ (b) and the filtered basin function with 300km Gaussian (c) and DDK2 (d) separately.

$$S_L = \frac{1}{R_0} \int_{(\Omega - R)} S_{leak} \hat{B} d\Omega. \tag{2.28}$$

3) The multiplicative correction approach assumes that water is uniformly distributed in a basin (Velicogna & Wahr, 2006; Longuevergne *et al.*, 2010), and obtains the reconstructed TWS by rescaling the leakage corrected TWS:

$$S_0 = (\hat{S}_0 - S_L) \cdot K, \tag{2.29}$$

by the multiplicative factor:

$$K = (\frac{1}{R_0} \int_R \hat{B} d\Omega)^{-1}. \tag{2.30}$$

Since we are estimating gridded TWS variations, the scaling-factor approach is used which can also be applied on a grid scale. Landerer & Swenson (2012) has provided 1 ° × 1° TWS estimates from GRACE using NOAH land model, running withing the Global Land-Data Assimilation System (GLDAS-NOAH). Long *et al.* (2015) applied the same approach using six land surface models (LSMs) and studied the impact of hydrological model uncertainties on the restored GRACE TWS anomalies. It is concluded that there are large impacts from the LSMs uncertainties especially in regions that are highly irrigated and/or arid and semiarid regions. The investigation, however, was only limited to the land surface models, most of which are close to each other in a way and did not consider the rescaling uncertainties within the TWS error estimation.

Chapter 3

GRACE-based gridded TWS variations and uncertainties

This chapter illustrates the specific way to estimate gridded TWS variations and uncertainties from GRACE gravity data in this thesis by applying the state-of-the-art post-processing procedure. Before that, post-processing methods are tested in a closed-loop environment by means of simulated GRACE-type gravity field time-series based on realistic orbits and instrument error assumptions as well as background error assumptions out of the updated ESA Earth System Model. Through the simulation test, it is intended to find the appropriate post-processing strategy applicable for the real GRACE data set. Furthermore, the method to estimate the approximated geocenter variations from GRACE monthly mean gravity fields alone will also be introduced. The degree-1 terms from the geocenter variations are then added to the GRACE gravity Stokes coefficients for further TWS estimation.

3.1 Investigating post-processing strategy on GRACE-like TWS variations for hydrological applications

As mentioned earlier, post-processing as filtering, de-striping and rescaling are needed to estimate TWS starting from the GRACE gravity stokes coefficients. For instance, different filtering methods from the basic Gaussian smoothing to the non-isotropic filtering (Swenson & Wahr, 2002; Han *et al.*, 2005; Chen *et al.*, 2007; Swenson & Wahr, 2006; Kusche, 2007) have been suggested to remove the high-frequency and correlated errors in GRACE gravity data. However, as an unique way to provide global large scale TWS variations, there is no independent data set that is comparable to validate the GRACE TWS estimates yet.

Simulation studies are generally performed before the launch of any satellite mission to find the balance between the scientific target and the overall mission expenses (Wiese *et al.*, 2012; Loomis *et al.*, 2012). These simulation results can also be applied to test the GRACE post-processing strategies in a closed-loop environment where the estimated TWS can be compared with the ones from the source model used in the initial orbit simulation. Flechtner *et al.* (2016) have performed such a full-scale simulation to investigate possible improvements of the laser ranging interferometer (LRI) technology demonstrator with respect to the basic microwave instrument (MWI) on GRACE Folllow-on (GRACE-FO). The MWI -based simulation data is applied here to investigate the optimal post-processing strategy for hydrometeorological applications. The residuals between recovered TWS variations and those from the true world source model provide another way to quantify the GRACE-based TWS uncertainties and to quantify how much we can reduce the GRACE/GRACE-FO TWS errors in future.

3.1.1 Simulation strategy

First, the simulation assumptions and strategy are introduced based on Flechtner *et al.* (2016). The simulation experiments covering five years from 01/2002 to 12/2006 are done based on a GRACE-like mission design. The initial altitude of the simulated orbits is 490 km which ends at 450 km after 5 years. This GRACE-like orbits also have an inclination of 89.0° and an eccentricity of 0.0015. An annual reset of the inter-satellite distance is applied to keep it between 170 km to 270 km. Besides, the solar activity values which are needed for both the non-gravitational forces and accelerometer data simulation are taken from the period 1995-1999 .

For the simulation, the first step is to forward the satellites' orbits and the true satellite-to-satellite tracking (SST) range-rate observations by propagating the satellites with a set of force models and initial conditions. The same software package is applied as for real GRACE data analysis: GFZ's Earth Parameter and Orbit System (EPOS). The background models used here for simulation is composed of a static gravity model up to degree and order 100 (EIGEN-GL04C; Förste *et al.*, 2008), 8 main constituents of the ocean tides from EOT08a (Savcenko & Bosch, 2008)), Sun and Moon ephemerides (DE405; Standish, 1998), and non-tidal mass variations in atmosphere, oceans, hydrology, ice, and solid Earth from the updated ESA Earth system model (Dobslaw *et al.*, 2015; Bergmann-Wolf *et al.*, 2014a). For the non-gravitational forces, atmospheric drag (Hedin, 1987), solar radiation, and Earth albedo (Knocke *et al.*, 1988) are modeled and transformed to acceleromenter data (ACC). Then 5-seconds orbits and SST observations are simulated.

In the next step, colored noise is added to the observations. The 5-seconds SST microwave instrument (MWI) noises are modeled as amplitude spectral density (ASD) (Elsaka *et al.*, 2014). The range-rate errors resulting from the accelerometry are derived from the inverse fast Fourier transform over the spectral density of GRACE-like accelerometer errors following the same procedure as Loomis *et al.* (2012).

The last step is to recover the gravity field from the simulated observations and added colored noises in the first two steps. In order to simulate an error level that is as realistic as possible, the background models are substituted here. The static gravity model and ocean tides are replaced by EGM96 (Lemoine *et al.*, 1997) and GOT4.7 separately. For the non-tidal atmospheric and oceanic mass variability, the realistically perturbed de-aliasing model from the updated ESA Earth system model is employed (Dobslaw *et al.*, 2016). Finally the 60 monthly gravity field models, expressed in terms of fully normalized SH coefficients up to degree and order 100 are obtained by performing again with the EPOS software.

3.1.2 The updated ESA Earth system model

As mentioned above, a model of the non-tidal mass variations at the Earth's surface is needed for the orbit computation. The updated ESA Earth system model is a new synthetic model which includes the mass variations in atmosphere, oceans, hydrology, ice, and solid Earth (AOHIS). Both the individual component and the sum of it are provided as spherical harmonic coefficients up to degree and order 180 with a temporal resolution of 6 hours covering a period from 1995 to 2006. The atmospheric component is represented by the latest ECMWF re-analysis data, ERA-Interim (Dee *et al.*, 2011). A modified IB correction has been applied to remove the correlation between the atmospheric and oceanic component of ocean bottom pressure (Dobslaw *et al.*, 2016). The ice component in Greenland and Antarctica is modeled with the regional climate model RACMO2 (Ettema *et al.*, 2009), while the low-frequency ice mass balance particularly determined by the ice dynamics, are taken from the re-synthesized ESM of Gruber *et al.* (2011). Ice mass balance for mountain glaciers and isolated ice-caps is included as linear trends for four groups of glaciers in Alaska, the European Alps, the Karakoram Mountain Range, and the Himalaya (Bergmann-Wolf, 2015). The Solid Earth component which includes three different signals: 1) GIA secular trends, 2) co-seismic deformations due to the Sumatra-Andaman earthquake between 00:00 and 6:00 on December 26th and (3) linearly increasing post-seismic deformation over one year, is also from the ESM of Gruber *et al.* (2011). The ocean component is composed of two models: (1) the Ocean Model for Circulation and Tides (OMCT; Thomas *et al.*, 2001) and (2) the STORM experiment from the MPIOM model (von Storch *et al.*, 2012). As OMCT can not

resolve the small scale variability, it mainly contributes to the lower spherical harmonics up to degree and order 60, whereas MPIOM contributes with the small-scale features from degree and order 61 to 180. The hydrology component is taken from the Land Surface Discharge Model (LSDM; Dill, 2008) -which will be introduced in detail in Section 4.1)- forced with the atmospheric fresh water and energy fluxes from ERA-Interim.

3.1.3 GRACE-like TWS estimation

Starting from this simulated gravity data, the TWS variations are estimated by applying different post-processing strategies. The SH coefficients are firstly truncated to degree and order 90 to be consistent with the real GFZ GRACE data. Then a multi-year mean covering the whole five year period is removed to obtain anomalies. The degree-1 terms are not considered here, and the C_{20} is substituted by the suggested GRACE TN07 values. Similar with GRACE, the GRACE-like monthly solutions are also contaminated by satellite errors mainly in the short wavelength (high degree) Stokes coefficients, and also aliasing errors due to the under-sampling of the short-term geophysical signals. The approximate decorrelation and non-isotropic smoothing method (DDK) introduced by Kusche (2007) is used here to remove such errors. Smoothing in space domain is enabled through a tuning parameter of the signal covariance matrix. Three different versions of DDK filter (DDK1, DDK2 and DDK3) that are approximately corresponding to an isotropic Gaussian radius of 530 km, 340 km and 240km separately (see Table A.1) are applied. The filtered spherical harmonic coefficients are synthesized into mass anomalies on the $1\,^{\circ} \times 1^{\circ}$ grids following the conventions of Wahr *et al.* (1998). Filtering and smoothing applied to the GRACE data will cause the basin-averaged TWS to deviate from the true value due to reduction of variance and incorporation of TWS variations from regions outside the basins of interest. In order to account for this effect, local re-scaling factors are introduced. The re-scaling factors are estimated from a small ensemble of four global hydrological models (GLDAS, WGHM, JSBACH, and MPI-HM). The median re-scaling factor is also calculated from these models to make it less affected by deficiencies in a certain model. Detailed introduction of these models are shown in Section 4.1 and the way to estimate the re-scaling factors will be illustrated in the Section 3.3. Since the main focus in this work is on the land water, the signals from ice and solid earth are removed over the same period. The gridded TWS variations are then averaged over the 50 largest basins.

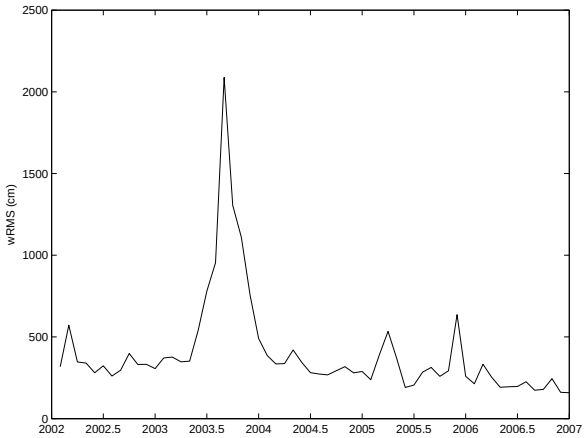

Figure 3.1: Global wRMS of unfiltered gridded GRACE-like TWS variations over the land (excluding Greenland and Antarctica).

3.1.4 Post-processing strategy investigation in a closed-loop environment

The estimated TWS variations are compared with the H component of the time variable AOHIS a priori model used for the simulation, which can be assumed as the "truth". To get a first impression on the quality of the GRACE-like TWS, global wRMS of the unfiltered GRACE-like TWS over the continent (excluding Greenland and Antarctica) are shown in terms of EWH. It is seen in Fig. 3.1 that there are four peaks in 02/2002, 08/2003, 03/2005 and 11/2005, which are due to the imperfect ground track coverage.

The differences between GRACE-like TWS and the variations from the H components are calculated over the continent. The time series of wRMS of the differences are shown and then we analyze the RMS of the differences (RMSD) covering the whole five year period in space domain. The RMSD comparison is shown for both grids on land and also the individual targeted basins. The performance of the different post-processing strategy is tested by comparing the minimum, maximum and wRMS of the RMSD.

3.1.4.1 Filtering

First, the appropriate DDK filter for the GRACE-like monthly gravity data is investigated. The wRMS time series of the differences between the GRACE-like TWS filtered with DDK1, DDK2 and DDK3 separately without re-scaling and the H component are shown in Fig. 3.2.

Generally, the application of DDK3 results in smaller wRMS compared to DDK1 and DDK2, except in those much noisier months as shown in Fig. 3.2, while the DDK1 performs just the

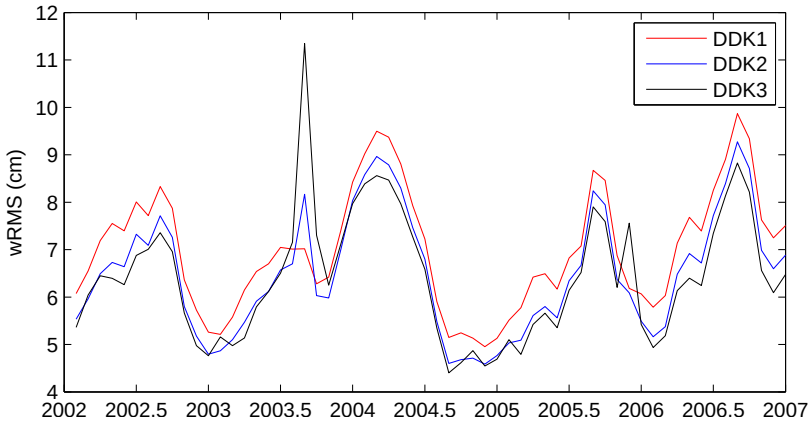

Figure 3.2: Global wRMS time series of the differences between the H component and the GRACE-like TWS variations filtered with DDK1, DDK2, DDK3 without re-scaling.

opposite. DDK2 lies in between and more close to DDK3, but does not show those peaks as DDK3. Fig. 3.3 exhibits the RMSD for the 50 chosen basins, where minimum, maximum and wRMS of the global values are also shown.

The "best" filter differs from basin to basin, depending on the basin size, shape and location, as indicated by Werth *et al.* (2009) when evaluating several filter methods for hydrological applications on a number of large river basins. For all the basins considered, DDK2 performs slightly better here, unlike the comparison on grids, as the basin averaging itself suppresses the noises at high degrees to some extend. Thus, DDK2 is chosen for the following calculation.

3.1.4.2 Rescaling

Gridded scaling factors estimated from a numerical model can be applied to compensate for the damping and leakage effect caused by the smoothing and filtering. However, the estimated scaling factors are also largely affected by the characteristics of the simulated water distribution in a larger area around the region-of-interest. Thus, uncertainties of the numerical model in terms of the model structure, parameter values and meteorological forcing will affect the estimated scaling factors as well. A small ensemble of 4 different numerical models is applied to calculate four sets of scaling factors and also the median values following the method as introduced in section 2.3.5. Then these scaling factors are applied separately on the DDK2 filtered TWS variations. Besides, scaling factors calculated from the H component which is assumed as the "truth", are also used for the re-scaled TWS estimation as a reference. Global wRMS time series of the differences are firstly shown in Fig. 3.4. The TWS variations rescaled by the median scaling factors show the smaller differences with the H component

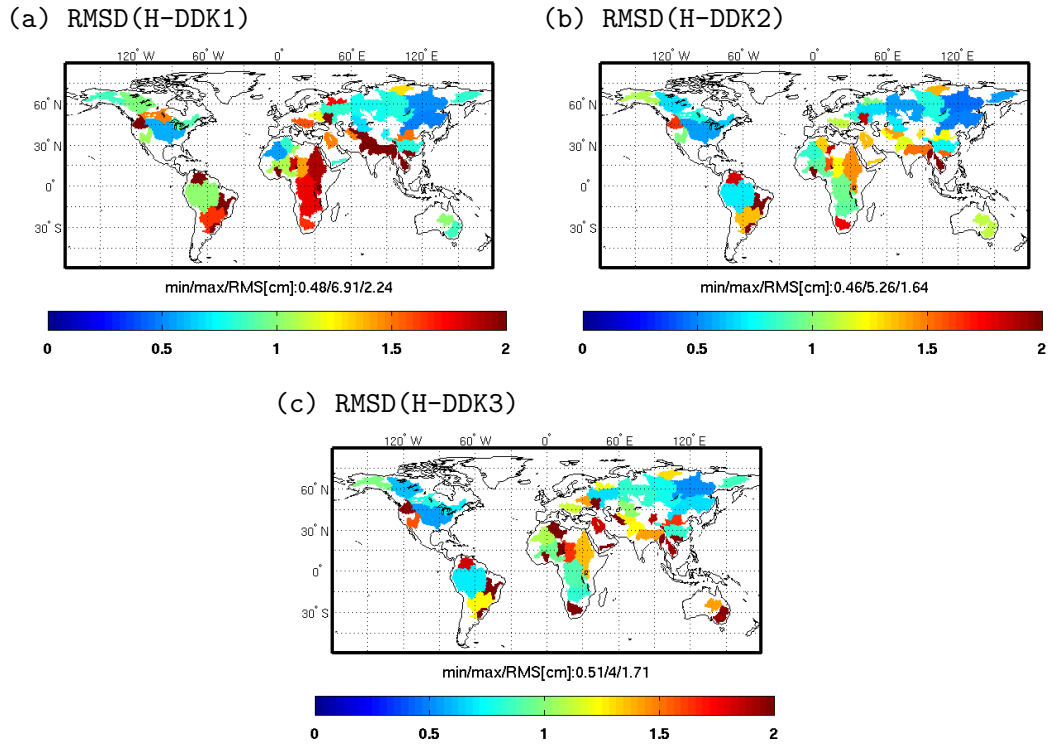

Figure 3.3: The RMS of the differences (RMSD) between the basin-averaged TWS from the H component and from the GRACE estimates filtered with DDK1, DDK2 and DDK3.

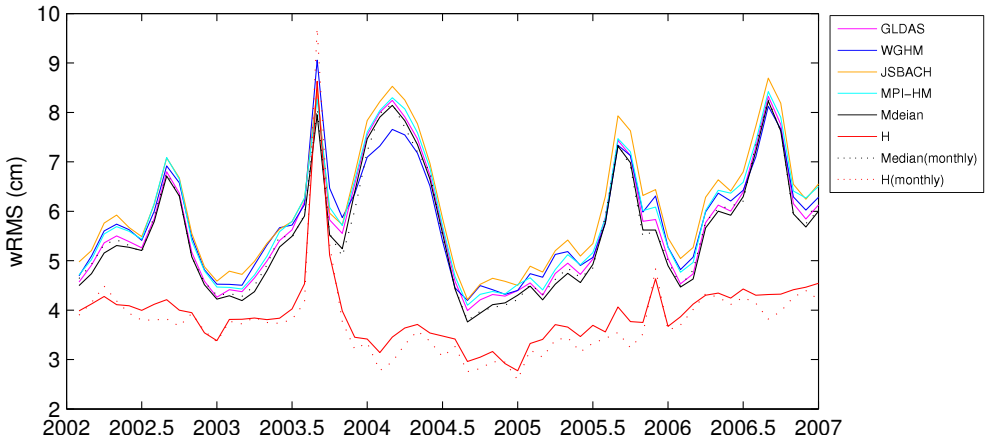

Figure 3.4: Global wRMS time series of the differences between the H component and the GRACE-like TWS variations filtered with DDK2 and rescaled with the scaling factors from GLDAS, WGHM, JSBACH, MPI-HM, median and the H component.

than those rescaled by an individual model except the time period from 01/2004 to 06/2004, indicating that the median of a small ensemble does make the scaling factors less affected by deficiencies in a certain model. However, there is still a gap with the result rescaled from the true hydrology signal (H) in terms of the wRMS of the differences. Further research efforts should be made to narrow this gap, which will significantly reduce the TWS uncertainties.

The gridded RMSD between model rescaled TWS and the H component are compared in Fig. 3.5. The patterns of the RMSD from the H component and TWS rescaled from different models are similar, with some differences at the local scale. The large RMSD mainly occurs at isolated high-variability regions close to the coast, at transition areas between the mountain and plains, and along spatially concentrated surface water bodies with high storage variability. The values of global wRMS over the continent are slightly different. Among the four models, although TWS rescaled from JSBACH exhibits the smallest minimum RMSD, it also shows the largest maximum and wRMS values; the smallest maximum value is from WGHM while the smallest wRMS is from GLDAS. TWS rescaled from the median exhibits the smallest wRMS value compared to those rescaled with individual model.

The RMSD for 50 chosen basins are shown as well in Fig. 3.6. The overall performance of the model rescaled TWS is different at basin scale from at grids. Here, MPI-HM shows the smallest minimum RMSD, but the largest maximum and wRMS values, while GLDAS has the smallest maximum RMSD and WGHM performs best in terms of wRMS. Median rescaled TWS shows the smallest wRMS of RMSD compared to the other four models, although not the smallest minimum and maximum RMSD. Still, the H rescaled TWS shows much better agreement with the "truth" (H). One large difference of the RMSD from median and H takes place in Nile basin. It is found that TWS variations at Nile basin is poorly simulated by the LSDM model which has been applied for the H component. The abnormal inter-annual variations in LSDM make the scaling factors quite different from the ones from other models. This confirms the earlier assumption that the scaling factors can be largely affected by numerical models uncertainties. The re-scaling errors should therefore not be neglected for the total TWS error estimation. Since re-scaling factors could vary temporarily because TWS changes spatially and temporally, we also calculate the time variable scaling factors for each month over the whole five year. However, only those rescaled with H monthly scaling factors show better results than the single scaling factor in terms of both gridded and basin-averaged RMSD, while for those rescaled from each individual model or median, the results are not improved (Table 3.1). Fig. 3.2 shows that TWS rescaled with median monthly scaling factors are quite close to the median scaling factors and the performance of H monthly scaling factors varies with time. This indicates that time variable scaling factors are very sensitive to the model uncertainties.

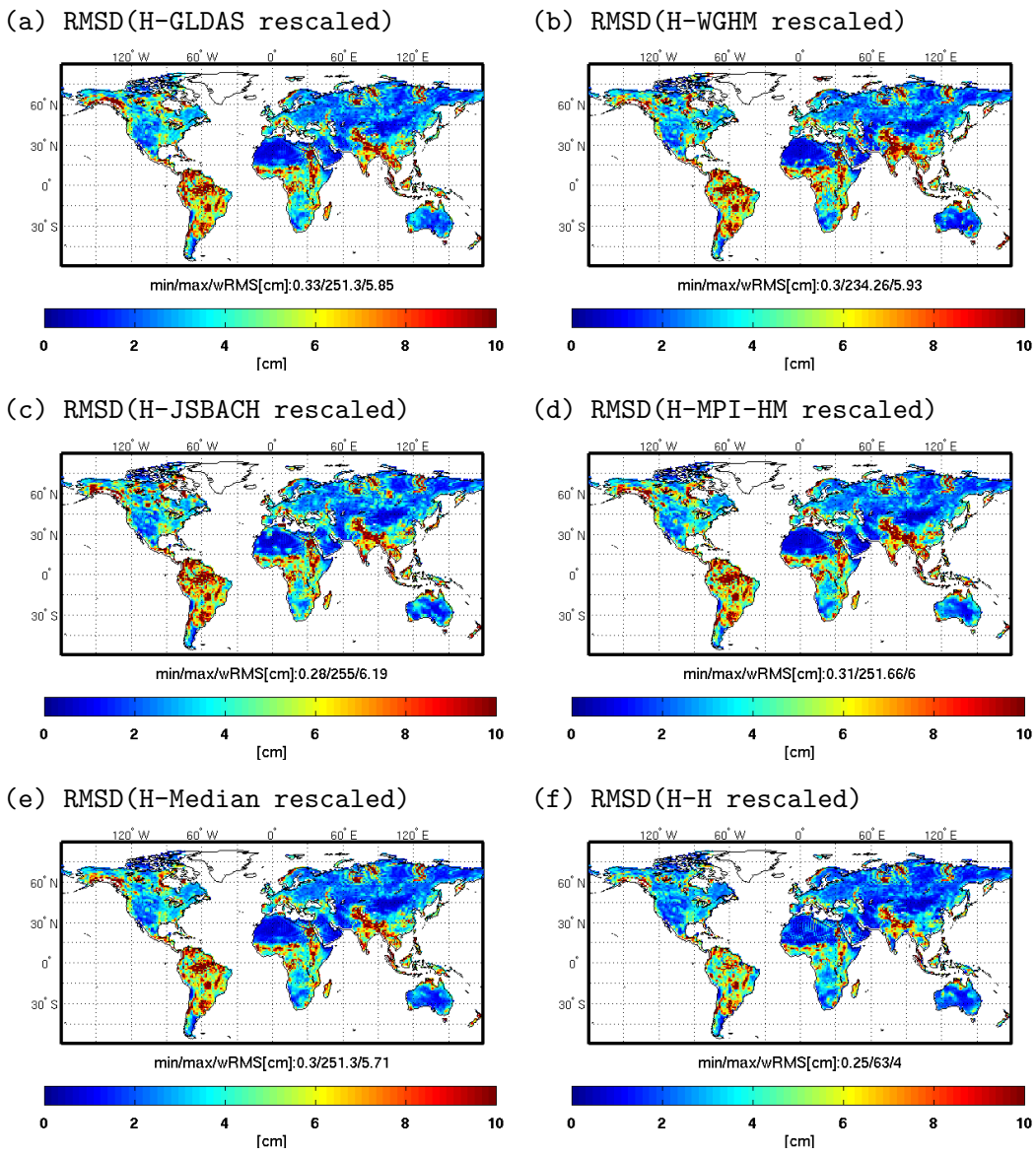

Figure 3.5: The RMS of the differences (RMSD) between H component and the GRACE-like TWS variations rescaled from GLDAS, WGHM, JSBACH, MPI-HM, median and H component over land grids.

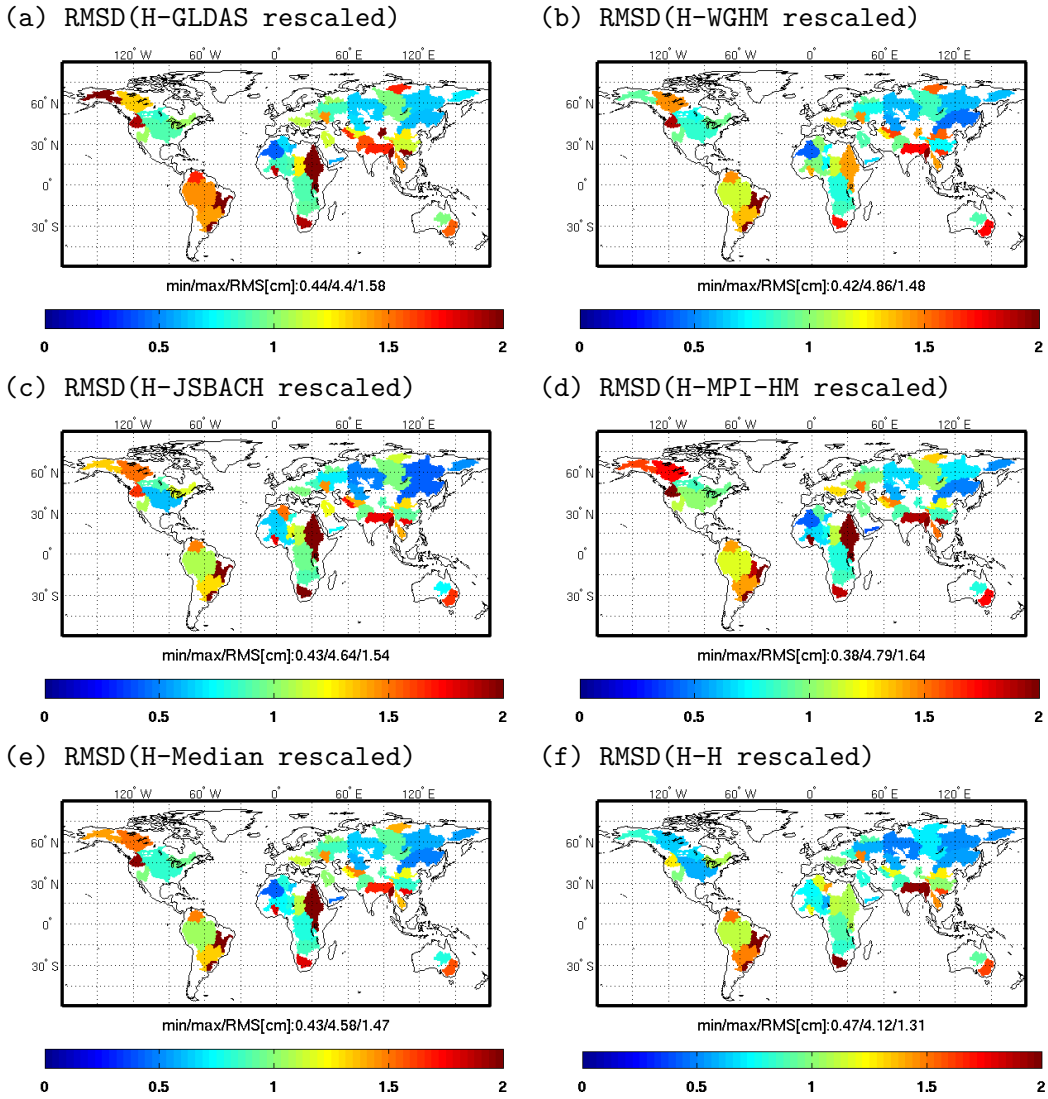

Figure 3.6: The RMS of the differences (RMSD) between the H component and the GRACE-like TWS variations rescaled from GLDAS, WGHM, JSBACH, MPI-HM, median and the H component for 50 specific basins.

Table 3.1: Minimum, Maximum and wRMS values (cm) of the RMS of the differences (RMSD) between the H component and rescaled GRACE-like TWS from monthly scaling factors for both grids on land and the individual targeted basins.

Model	spatial scale	Minimum	Maximum	wRMS
GLDAS	grids	0.47	252.27	6.3
	basins	0.47	8.28	1.97
WGHM	grids	0.5	234.62	6.42
	basins	0.45	4.87	1.6
JSBACH	grids	0.38	255.73	6.58
	basins	0.54	4.64	1.64
MPI-HM	grids	0.42	270.43	6.49
	basins	0.44	5.22	1.74
Median	grids	0.45	246.55	5.74
	basins	0.44	4.52	1.46
H	grids	0.28	78.65	3.89
	basins	0.51	3.39	1.26

3.1.4.3 GRACE-like TWS uncertainty estimates

The TWS uncertainties can also be estimated from the the contributions of measurement errors, leakage errors, and re-scaling errors following the method from Landerer & Swenson (2012) as demonstrated for the latest GFZ GRACE release by Zhang *et al.* (2016). The measurement error is estimated by error propagation from the "calibrated errors", which is then multiplied with the gridded re-scaling factors. The leakage error is calculated as the RMS differences between the original and filtered signals from H components, rescaled by the ratio of RMS of the TWS from GRACE and H component. The re-scaling error is then calculated by multiplying the RMS of the GRACE signals with the difference between the re-scaling factor from the H component and the median value. The total error at each grid point is subsequently taken as the sum of the measurement error, leakage error, and re-scaling error in quadrature. Details on the method will also be given in Section 3.3. The basin-averaged water storage errors are then estimated considering the correlations between the grids using the squared exponential covariance function as proposed by Landerer & Swenson (2012). The same de-correlation length scales as in Zhang *et al.* (2016) are applied, which are obtained by fitting the error budget from the gridded data set to the ones obtained from estimating errors directly at basin-scale level. The error estimates (Fig. 3.7) are consistent with the RMSD between GRACE-like TWS and the H component (Fig. 3.6 e) and generally larger, implying that those are rather conservative error estimates.

Based on the above investigation, we would suggest to apply DDK2 filter and median scaling factors from an ensemble of several models on real GRACE gravity data if no additional

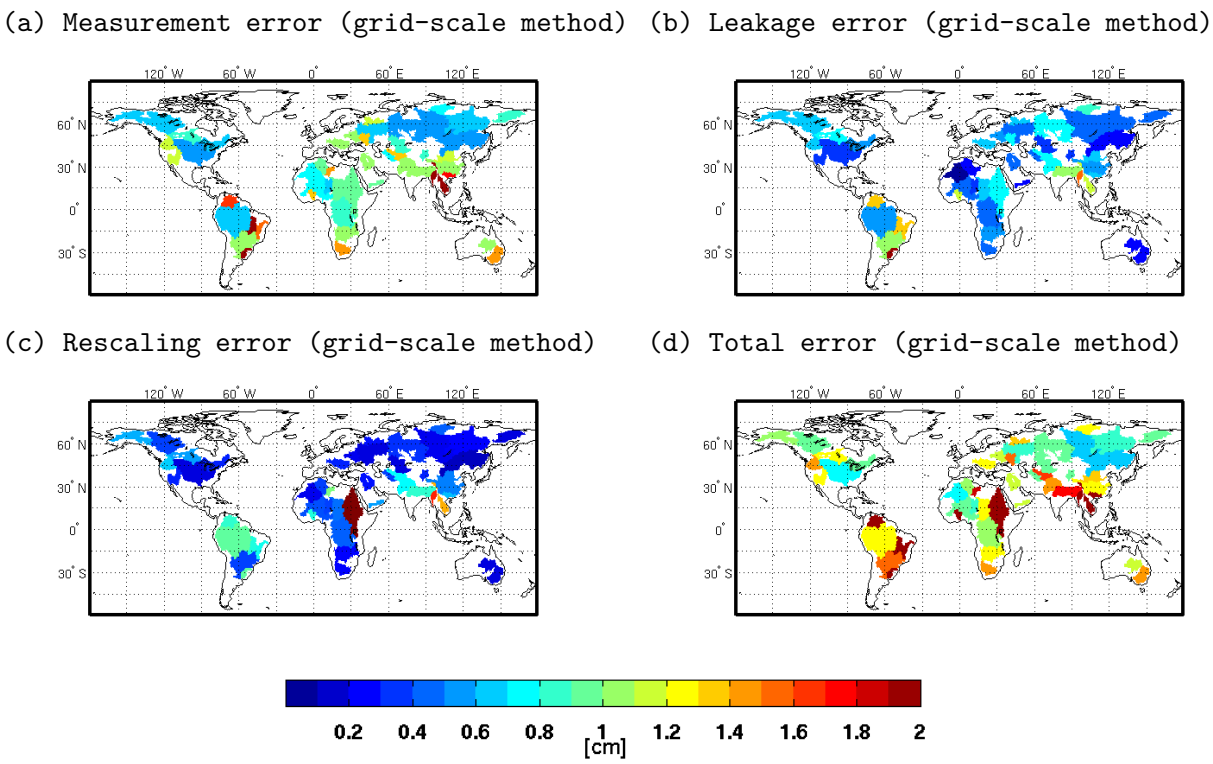

Figure 3.7: Estimates of GRACE-like TWS errors for 40 largest discharge basins derived from 1° grid point estimates: measurement errors (a), leakage errors (b), re-scaling errors (c), and total errors (d).

information is available. The estimated total error can serve as an upper bound for the GRACE TWS uncertainties.

3.2 Geocenter motion approximation from GRACE

Based on: Bergmann-Wolf, I., **Zhang, L.** and Dobslaw, H. (2014). Global eustatic sea-level variations for the approximation of geocenter motion from GRACE, J. Geod. Sci., 4, 37 – 48, doi:10.2478/jogs-2014-0006.

The global gravity field solutions are typically calculated in a center of mass (CM) frame, so the degree-1 harmonics are not measurable by the GRACE mission. When compared with other independent data sets which are defined in the center of figure (CF) (or center of solid earth (CE)) frame, for instance the hydrological observations and models, the degree-1 term should also be considered. Swenson *et al.* (2008) proposed a method that infers global degree-1 terms from a-priori knowledge of mass distribution at a sufficiently large fraction of the Earth's surface (for instance, the global ocean) in combination with Stokes coefficients of the higher degree and order as derived from GRACE observations. Thus a model of the global eustatic sea-level variations is typically needed. Since eustatic sea-level variability can be assumed to be globally homogeneous when loading and self-attraction effects are not considered (Dobslaw & Thomas, 2007; Kuhlmann *et al.*, 2011) and it is included in the GRACE monthly mean gravity fields (Chambers *et al.*, 2004), it can also be derived empirically from the GRACE data set. Based on the Swenson *et al.* (2008), the method to estimate the eustatic sea-level variations and the geocenter motion in an iterative way is illustrated.

3.2.1 Review of Swenson's method

The method of Swenson *et al.* (2008) is based on the relationship between the global degree-1 coefficients and the oceanic components. Mass anomalies in a certain region (here global ocean is taken) can be obtained from the knowledge of the global Stokes coefficients and the area extent of that region. Subsequently the global degree-1 coefficients can be expressed by a linear combination of the oceanic degree-1 components which can be provided by other auxiliary sources and the higher degree spherical harmonic coefficients from GRACE.

The surface mass anomalies in a certain region (here global ocean is taken) can be obtained from the knowledge of the global mass anomalies and the area extent of that region:

$$\sigma^{ocean}(\vartheta, \lambda) = \sigma(\vartheta, \lambda) B(\vartheta, \lambda) \tag{3.1}$$

where B is the ocean function:

$$B(\vartheta,\lambda)=\begin{cases}1 & if \quad ocean\\ 0 & if \quad land\end{cases}, \tag{3.2}$$

and the global mass anomalies can be estimated from the spherical harmonic coefficients as shown in 2.13.

The C_{lm}^{ocean} (denoted as ocean coefficients) can then be expressed as:

$$\Delta C_{lm}^{ocean}=\frac{1}{4\pi}\int \tilde{P}_{lm}(\cos\vartheta)\cos m\lambda B(\vartheta,\lambda)\sigma(\vartheta,\lambda)d\vartheta d\lambda, \tag{3.3}$$

and when the surface mass $\sigma(\vartheta,\lambda)$ is expressed in terms of the surface density coefficients, it is denoted as:

$$\begin{aligned}\Delta C_{lm}^{ocean}=&\frac{1}{4\pi}\int \tilde{P}_{lm}(\cos\vartheta)\cos m\lambda B(\vartheta,\lambda)\\ &\times\sum_{l'=0}^{\infty}\sum_{m'=0}^{l'}\tilde{P}_{l'm'}(\cos\vartheta)(\Delta C_{l'm'}\cos m'\lambda+\Delta S_{l'm'}\sin m'\lambda)d\vartheta d\lambda\end{aligned} \tag{3.4}$$

If we take the degree 1 global coefficients out and move it to the left of the equation, it can be expressed by the oceanic coefficients and the remaining degree spherical harmonic coefficients. For instance, for C_{10}:

$$\Delta C_{10}=\frac{4\pi C_{10}^{ocean}-\int \tilde{P}_{10}(\cos\vartheta)B(\vartheta,\lambda)\sum_{l=0}^{\infty}\sum_{m=0}^{l}\tilde{P}_{lm}(\cos\vartheta)(\Delta C_{lm}\cos m\lambda+\Delta S_{lm}\sin m\lambda)d\vartheta d\lambda}{\int \tilde{P}_{10}(\cos\vartheta)B(\vartheta,\lambda)\tilde{P}_{10}(\cos\vartheta)d\vartheta d\lambda} \tag{3.5}$$

To consider the other two degree 1 terms and make the equations more concise, we can re-arrange the equations into a single rank 3 matrix equation:

$$\begin{bmatrix}\Delta C_{10}^{ocean}\\ \Delta C_{11}^{ocean}\\ \Delta S_{11}^{ocean}\end{bmatrix}=\begin{bmatrix}I_{10C}^{10C} & I_{11C}^{10C} & I_{11S}^{10C}\\ I_{10C}^{11C} & I_{11C}^{11C} & I_{11S}^{11C}\\ I_{10C}^{11S} & I_{11C}^{11S} & I_{11S}^{11S}\end{bmatrix}\begin{bmatrix}\Delta C_{10}\\ \Delta C_{11}\\ \Delta S_{11}\end{bmatrix}+\begin{bmatrix}G_{10C}\\ G_{11C}\\ G_{11S}\end{bmatrix}, \tag{3.6}$$

where for instance, I_{10C}^{11C} is shown as:

$$I_{10C}^{11C}=\frac{1}{4\pi}\int \tilde{P}_{11}(\cos\vartheta)B(\vartheta,\lambda)\tilde{P}_{10}(\cos\vartheta)d\vartheta d\lambda, \tag{3.7}$$

where the superscript indicates the spherical harmonic to the left of $B(\vartheta, \lambda)$ and the subscript the one to the right, and

$$G_{11C} = \frac{1}{4\pi} \int \tilde{P}_{11}(\cos\vartheta) B(\vartheta, \lambda) \times \sum_{l=0}^{\infty} \sum_{m=0}^{l} \tilde{P}_{lm}(\cos\vartheta)(\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda) d\vartheta d\lambda. \tag{3.8}$$

The degree-1 terms can then be estimated by matrix inversion with the oceanic degree-1 terms from auxiliary sources and the higher degree coefficients (l>2) from GRACE. Since the eustatic signal from the monthly water transport between the land and ocean has been intentionally excluded from the de-aliasing model (Flechtner & Dobslaw, 2013), it should also be added to the oceanic degree 1 terms.

3.2.2 Global eustatic mass variations from GRACE

The eustatic sea-level variability can be estimated from the GRACE gravity solutions Chambers *et al.* (2004). Here, the GRACE RL05 gravity field solutions from GFZ up to d/o 90 are used. Since degree-1 terms are also needed for this estimation, an annual sinusoid of degree-1 coefficients is taken from Eanes (2000). The ocean mass variability is then estimated with similar approach as basin-average on land shown in Eq. 2.25. The ocean area, however, is taken as only the domain 300km away from the coasts to minimize continental leakage. Furthermore, 300km Gaussian filter is applied to reduce the systematic errors. Since the filtering can cause attenuation of the signal, a rescaling factor is also applied.

3.2.3 Degree-1 coefficients estimation from GRACE

For the higher degree spherical harmonic coefficients (l>2), we also used the GRACE RL05 gravity field solutions from GFZ. The contribution to the GSM coefficients from glacial isostatic adjustment (GIA) are estimated from the model of Paulson *et al.* (2007). The GAD products from the GRACE project, which are calculated from the Ocean Model for Circulation and Tides (OMCT; Thomas *et al.*, 2001), are used to represent the oceanic components. The eustatic degree-1 components as estimated above are also considered. The degree-1 terms (shown in Fig. 3.8) are very similar to the results from the GRACE TELLUS website based on CSR RL5 solutions. As a priori information of the degree-1 terms are needed for the eustatic sea-level variability estimation, it can now be replaced by this newly derived global degree-1 coefficients. Both the degree-1 coefficients and the eustatic sea-level variability can now be calculated in a iterative way (Bergmann-Wolf *et al.*, 2014b). The results converge

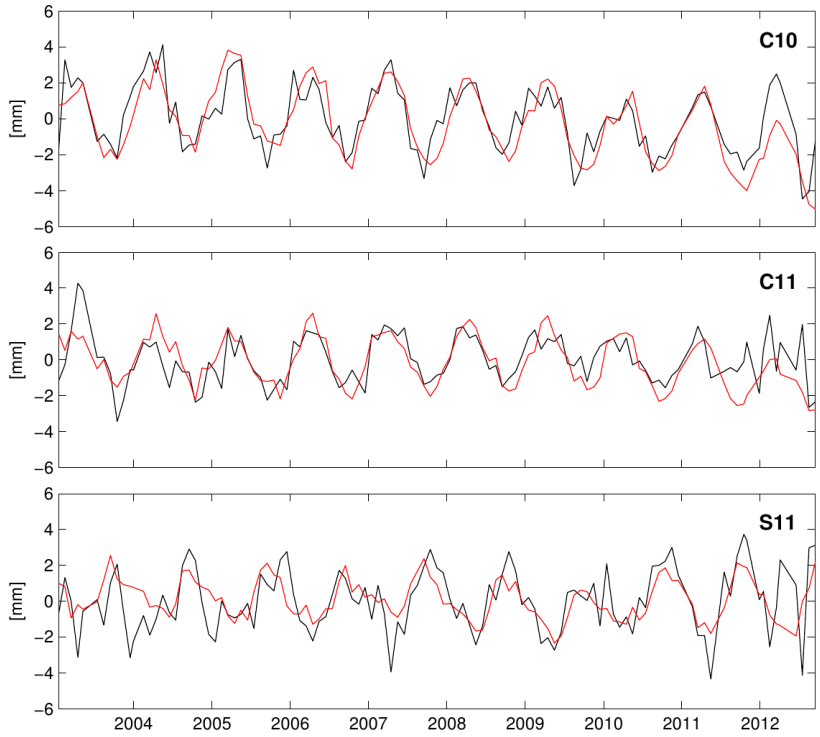

Figure 3.8: Degree-1 terms variation time series from GFZ RL05 (black) with global ocean mass variations estimated with 300 km Gaussian filter, compared with the ones provided by the GRACE TELLUS webpage (http://grace.jpl.nasa.gov/data/degree1/) for the CSR RL05 solutions (red) (Bergmann-Wolf *et al.*, 2014b).

after about 5 iteration with decreasing increments. Besides, even when the initial information of degree-1 terms are not included for the eustatic sea-level calculation (experiment 1) and the ocean degree-1 terms are set to zero for geocenter estimation (experiment 2), after iterating for 5 times, the final results converge and agree well with previous calculation (Fig. 3.9). This indicates that the degree-1 terms can be calculated from GRACE monthly products without additional model information. The degree-1 terms estimated based on the same approach are then added to the corresponding GRACE monthly gravity solutions for further calculation.

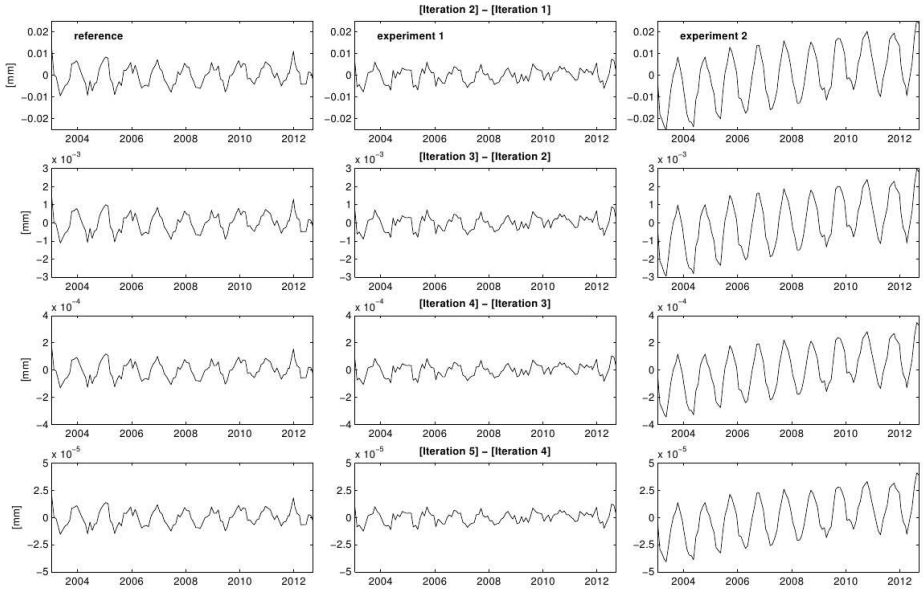

Figure 3.9: Increments in eustatic sea-level from the iterative determination of global degree-1 coefficients and a globally homogeneous eustatic sea-level variability model: reference is calculated from global degree-1 coefficients given by Eanes (2000) (left column), experiment 1 starting from from zero global degree-1 coefficients (middle), and experiment 2 is from zero eustatic sea-level anomaly (right) (Bergmann-Wolf *et al.*, 2014b).

3.3 Gridded TWS variations from GRACE

Published as: Zhang, L., Dobslaw, H., and Thomas, M. (2016): Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications, Geophys J Int., 206 (1): 368 – 378. doi: 10.1093/gji/ggw153.

3.3.1 Introduction

Satellite observations of temporal variations of the Earth's gravity field from the Gravity Recovery and Climate Experiment (GRACE) satellite mission (Tapley *et al.*, 2004b) offer a new possibility to monitor a wide range of Earth system dynamics that are related to large-scale mass re-distribution. Over the continents, the mission is in particular sensitive to several aspects of the terrestrial branch of the global water cycle: it allows measurements of the mass balance of continental ice-sheets and glaciers (Jacob *et al.*, 2012), deep soil moisture variability and its consequences for drought and flood potentials (Houborg *et al.*, 2012), as well as groundwater depletion arising from growing water demands for irrigation and human consumption in agricultural regions world-wide (Voss *et al.*, 2013).

The anomaly of terrestrial water storage (TWS) with respect to a long-term average is the most direct hydrological quantity obtainable from the GRACE science products. TWS is understood here as the sum of all storage compartments of water at and underneath the land surface. This includes soil moisture; the water content in snow-pack and land ice; ground water in shallow and deep aquifers; canopy water; and also the content of surface water bodies as rivers, lakes, and occasionally flooded wetlands. As a new observable that is available globally from space-based instruments, GRACE-based TWS serves an important role in assessing the closure of the terrestrial water balance at global and regional scales and allows for a new way to assess and even improve the quality of hydrological model simulations (Syed *et al.*, 2008a; Frappart *et al.*, 2013; Eicker *et al.*, 2014).

Due to its observing principle, GRACE data is highly accurate at hemispheric spatial scales but provides no information on TWS variability at spatial scales smaller than a few hundred km. GRACE monthly mean gravity fields and their associated uncertainties are typically provided in terms of global spherical harmonic coefficients – a mathematical representation that substantially hampers the application of GRACE mission results in non-geodetic branches of the physical geo-sciences. In order to overcome this limitation, we describe in this paper a globally gridded data-set of TWS anomalies at monthly resolution that is corrected for known systematic errors and other deficits by applying state-of-the-art post-processing methods. To allow for a subsequent application of the data-set in model validation

efforts or even data assimilation experiments, globally gridded observational uncertainties are derived as well. Similar gridded TWS products are also available at the Tellus website (http://grace.jpl.nasa.gov/data/gracemonthlymassgridesland/) which are calculated using a different processing scheme as described by Landerer & Swenson (2012). For simplicity, we use the term Tellus to refer to this data-set throughout the manuscript.

The structure of the paper is as follows: The post-processing procedure to derive globally gridded TWS anomalies out of the GRACE monthly mean gravity fields will be introduced in detail in Sect. 3.3.2. Subsequently, we introduce our method to correct for the leakage and bias using median scaling factors estimated from a small ensemble of five hydrological model simulations in Sect. 3.3.3. In Sect. 3.3.4, we present our gridded error estimates, whose consistency with uncertainties directly obtained from the Stokes coefficients are demonstrated for the largest 50 basins. We then compare the basin-averaged TWS from our gridded TWS variations with those from Tellus for these basins in Sect. 3.3.5 to demonstrate the consistency and robustness of our method, and end with some concluding remarks on the general applicability of GRACE results for hydrometeorological applications in the final Sect. 6.

3.3.2 GRACE data-sets and post-processing

GRACE is a twin satellite mission of NASA and the German space agency DLR that was launched in March 2002 into a polar orbit at an initial altitude of only 450 km. By means of highly accurate range-rate measurements between the two space-crafts that follow each other with a typical separation distance of 250 km, the mission is able to map the Earth's gravity field at approximately monthly intervals on spatial scales of a few hundred km and larger. After removing short-term variability due to tides in solid Earth (Petit & Luzum, 2010), oceans (Savcenko & Bosch, 2012) and atmosphere (Biancale & Bode, 2006), as well as non-tidal variability in atmosphere and oceans (Dobslaw *et al.*, 2013) from the observations, the remaining gravity changes on monthly to inter-annual time scales mainly represent variations in terrestrial water storage. Since the mission inherently does not provide vertical resolution but is instead equally sensitive to mass variability at or beneath the surface, an unambiguous separation of individual storage compartments by means of GRACE data only is impossible.

We use the monthly GRACE release 05a Level-2 products expanded in spherical harmonics up to degree and order 90 – corresponding to a spatial wavelength of 220 km – from GFZ Potsdam (Dahle *et al.*, 2012), which show an overall improvement by a factor of two in terms of noise reduction compared to previous releases (Chambers & Bonin, 2012). The data can be visually explored and downloaded from the website of the International Centre for Global Earth Models (ICGEM) available at "icgem.gfz-potsdam.de/ICGEM". In addition to this routinely

updated standard GRACE solution, we further utilize in this thesis a recent GRACE release from the University of Graz (ITSG-Grace2014; Mayer-Gürr *et al.*, 2014), which additionally provides full variance-covariance matrices consistent with its unconstrained monthly gravity field solutions.

For the derivation of globally gridded TWS variations from GRACE gravity fields we essentially follow the strategy chosen by Bergmann & Dobslaw (2012), which is briefly recalled here. First, we add global degree-1 coefficients as derived by Bergmann-Wolf *et al.* (2014b) with the methodology of Swenson *et al.* (2008), replace the C_{20}-coefficients with estimates from Satellite Laser Ranging (Cheng *et al.*, 2011), and remove a multi-year average over the period of January 2004 to December 2009 which is consistent with Tellus to arrive at anomalies. Next, we apply the approximate de-correlation and non-isotropic smoothing method introduced by Kusche (2007) to remove correlated errors in north-south directions that are related to the anisotropic sensitivity of the track-aligned range-rate measurement system between the two satellites. Smoothing in space is enabled through the tuning parameter a=10^{13} of the signal covariance matrix, which is known also as DDK2 filter and approximately corresponds to an isotropic Gaussian filter with 680 km full width half maximum (Kusche *et al.*, 2009). The filtered spherical harmonic coefficients are finally synthesized on a 1° latitude-longitude grid following the conventions of Wahr *et al.* (1998). Since assuming that mass redistribution occurs at the surface of a sphere may introduce errors when spherical harmonic coefficients are expanded to higher degrees, we use a reference ellipsoid as defined by the IERS conventions which is a more suitable approximation of the Earth's shape. Note that using instead a spherical surface might only impose differences of up to 0.5 cm equivalent water height (e.w.h.) in particular at the higher latitudes.

3.3.3 Compensation of filter-induced signal attenuation

The gravitational field of the Earth is conservative in space and might be thus continued both upward to the satellite orbit and downward to the Earth's surface without loss of generality. The signal decay with height depends on the spatial extent of a gravity disturbance, so that large-scale anomalies generally cause stronger deviations of a spacecraft trajectory from its reference orbit. Errors of GRACE-based TWS variations therefore grow with increasing degree of the spherical harmonics expansion, and low-pass filtering in space or spectral domain is consequently required to reduce the contribution from highly uncertain smaller spatial scales when gridded estimates are to be calculated.

The process of truncation and filtering, however, typically diminishes amplitudes of highly localized signals, and increases spatial leakage of signals from neighboring regions (Werth

et al., 2009). In order to approximately account for such effects in the post-processing, Klees *et al.* (2007) suggested to apply local re-scaling coefficients derived from TWS predictions from a global numerical model. The scaling factors estimated from a numerical model are, however, strongly related to the characteristics of the simulated water distribution in a larger area around the point-of-interest. Thus, uncertainties of the numerical model in terms of the model structure, parameter values and and meteorological forcing will affect the estimated scaling factors as well. Recent inter-comparison studies of hydrological models demonstrated that there can be a large spread in model performance (see, e.g., Gudmundsson *et al.*, 2012) for various regions and frequencies. Thus, the sensitivity of the re-scaling factors to different models is investigated in this paper more closely by using a small ensemble of 5 different numerical model experiments, which includes land surface schemes of both low and intermediate complexity; global land assimilation systems; and conceptual hydrological models.

(1) LSDM forced with operational ECMWF analysis data.

(2) A single realization of GLDAS based on the Noah community land surface model downloaded from "mirador.gsfc.nasa.gov".

(3) WGHM of version 2.2 as described by Müller Schmied *et al.* (2014) and has been forced with atmospheric data specifically prepared for the Water and Global Change (WATCH) project, which is based on ERA-Interim re-analysis data (Dee *et al.*, 2011) and bias corrections for precipitation from the Climate Research Unit (CRU) station-based monthly climatologies (WFDEI_CRU; Weedon *et al.*, 2011).

(4) An un-coupled version of JSBACH that is also driven by daily the WFDEI_CRU atmospheric data.

(5) MPI-HM that is also integrated with WFDEI_CRU atmospheric forcing data.

Detailed information of the models can also be found in Section 4.1. Besides the apparent differences in model structure, parametrization, and atmospheric forcing, the model experiments considered here also differ with each other by what particular water storage compartments are actually included in TWS (Table 4.1). The results from all the model runs are aggregated into monthly averages and – where necessary – conservatively interpolated onto a regular latitude-longitude grid with a horizontal spacing of 1°. Since GRACE estimates over Greenland and Antarctica are dominated by surface mass balance and ice dynamics of the continental ice-sheets, both regions have been masked out of the model results together with the global oceans and will not be considered further in the reminder of this paper.

All model experiments and GRACE data are available to us over the period 2003 – 2012. Since local trends are usually poorly predicted by the models considered here, and since the

influence of GIA signals cannot be completely removed due to a lack of knowledge on both glacial ice load history and mantle viscosity (Steffen *et al.*, 2008), we concentrate on TWS variability on seasonal to inter-annual time-scales only. Linear trends are therefore estimated and subtracted from all data-sets considered.

Local re-scaling factors that compensate for filter-induced signal alteration have been obtained from all five model runs by calculating the average ratio between monthly TWS before and after filtering with DDK2 (Fig. 3.10). Re-scaling factors are almost zero for arid regions dominated by leakage-in from neighboring regions with stronger signal variability. But factors are also as large as three for (i) isolated high-variability regions close to the coast, which are affected by leakage-out due to the much weaker ocean bottom pressure variability near-by; at (ii) step gradients in orography that are responsible for locally intensified precipitation; or (iii) along spatially concentrated surface water bodies with high storage variability. By comparing re-scaling factors for the different models, we note moreover substantial discrepancies among the model runs, in particular at smaller spatial scales. There is, however, no prior knowledge of which model performs best at different areas of the world, so that we calculate a median value of the re-scaling coefficients from the five model runs at each grid point, which is finally applied to the GRACE data.

To further quantify the uncertainties in the re-scaling factors, we calculate root mean square (RMS) estimates of the difference between each re-scaling factor and the median value normalized by the median itself (Fig. 3.11). Those variation coefficients show similarities with the results from Long *et al.* (2015), which is quite opposite with the pattern of the signal variability of GRACE-based TWS (Fig. 3.12). Largest variability of the re-scaling factors occurs in rather dry areas, like North Africa, South Australia, Middle East, and Northwest China, which is partly explained by the small values of the median scaling factors in these regions. Areas dominated by surface water variability and affected by groundwater abstraction also show large spreads of the re-scaling factors.

We also look into the re-scaling coefficients estimated from different models at basin-scale level (Table 3.2). Despite of the differences indicated above, variation coefficients of the re-scaling factors are mostly below 0.15 when averaged over areas of several 10^6 km^2, thereby indicating the high consistency of the TWS simulated by all models in these large areas. Uncertainties grow when smaller basins with a lower GRACE signal-to-noise ratio (SNR) are considered. The SNR is calculated as the ratio of the RMS of the GRACE TWS time series and the total error estimated from the basin-scale method which is introduced in Sect. 3.3.4. Both the low GRACE SNR and the large spread of the scaling factors from hydrological models confirm the poor ability of GRACE and hydrological models to capture the TWS signal in overly

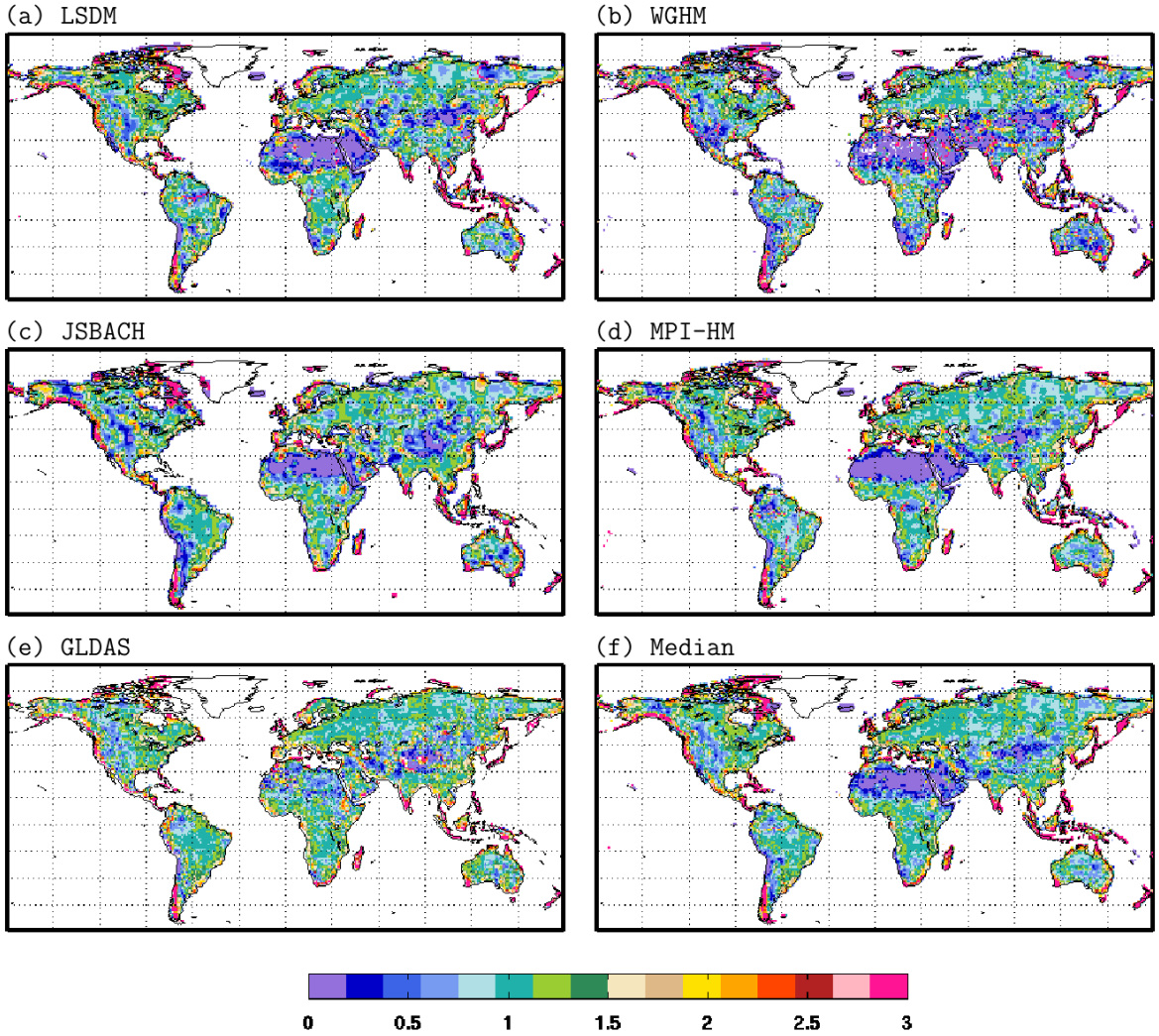

Figure 3.10: Re-scaling factors obtained as the average ratio of DDK2-filtered and unfiltered TWS time-series from five global land model experiments performed with LSDM (a), WGHM (b), JSBACH (c), MPI-HM (d), and GLDAS (e). In order to reduce the impact of individual model deficiencies, the median value (f) of all five simulations is applied in the GRACE post-processing.

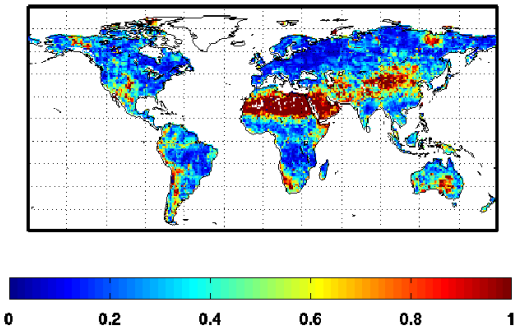

Figure 3.11: Variation coefficients for the re-scaling factors as obtained as the RMS between individual model-derived re-scaling factors and the median value normalized by the median.

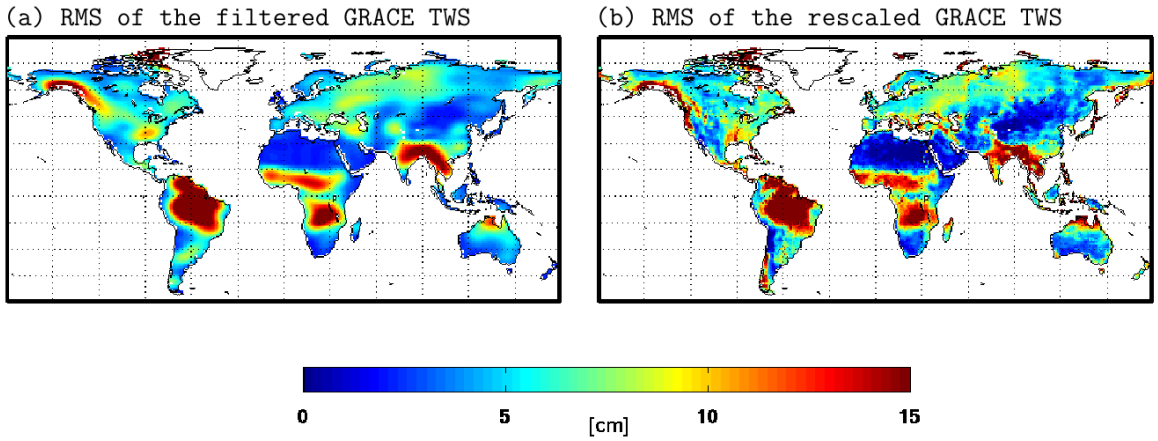

Figure 3.12: RMS variability of GRACE-based terrestrial water storage after applying the DDK2 filter (a), and after re-scaling of the filtered results (b) with spatially variable median re-scaling factors obtained from five different model data-sets.

dry areas (Gunkel & Lange, 2011). Care should also be taken in some humid areas where a large spread of the re-scaling factors is found, as, e.g., in the Yukon basin, where a substantial contribution of mountain glaciers on observed TWS variability can be expected which is rather poorly represented in all the models. Considering that we only focus on the seasonal and interannual signals, it is therefore not suitable to apply such scaling factors in areas which have large contribution from glaciers. Further, observed TWS in the Chang Jiang basin is affected by surface water variability, which is not represented properly in both GLDAS and JSBACH. In the Indus catchment, intensive irrigation takes place which uses both surface water and groundwater resources. However, only WGHM and JSBACH accommodate groundwater storage changes in their physical models. Nevertheless, the application of the median value makes the re-scaling factors less affected by such deficiencies in a single model and therefore contributes to the robustness of the GRACE post-processing methodology applied here.

3.3.4 TWS errors from GRACE

In order to provide a quantitative estimate for the uncertainties associated with GRACE-based TWS errors, we individually assess the contributions of measurement errors, leakage errors, and re-scaling errors as suggested by Landerer & Swenson (2012). We estimate the measurement error by error propagation from the "calibrated errors" provided together with the GRACE monthly-mean Stokes coefficients. The calibrated errors are further re-scaled to fit the non-seasonal GRACE residuals after subtracting a constant as well as annual and semiannual sinusoids following Wahr *et al.* (2006). Then gridded re-scaling factors are multiplied with the measurement errors to get the final measurement error distribution in the spatial domain. The leakage error compartment is calculated for all five model realizations according to

$$E_{leak} = RMS(\Delta S_o - k\Delta S_F)\frac{RMS_{GRACE}}{RMS_{Model}}, \tag{3.9}$$

where ΔS_o and ΔS_F are the original and filtered signals from the models, respectively, k is the re-scaling factor, and RMS_{GRACE} and RMS_{Model} are the root mean square (RMS) of the TWS from filtered GRACE and from one of the unfiltered model data-sets. For the re-scaling error, we make use of the scaling ratio for the individual model run shown in Fig. 3.10 by multiplying the RMS of the GRACE signals with the difference between each realization-based re-scaling factor with the median value. The total error at each grid point is subsequently taken as the sum of the measurement error, leakage error, and re-scaling error in quadrature.

While calculating these error compartments individually for each grid point, we obtain total errors of up to 10 cm in equivalent water height (e.w.h.; Fig. 3.13). The water storage variation estimated from GRACE is not a point measurement, but rather a regional spatial average. It therefore does not make sense to compare GRACE at 1° grid-scale level directly to any other data-set. We use, instead, the gridded error estimates as a starting point for deriving error estimates of arbitrarily shaped regional averages. Here, the shapes of the 50 largest river basins from Table 3.2 are chosen, but the methodology can be similarly applied to other areas as well, as e.g., climate or altitude zones associated with particular precipitation regimes, the spatial extent of a specific aquifer system, or an area particularly affected by land use and land cover changes. Since the error contributors of GRACE-based TWS at the grid scale are spatially correlated, the basin-averaged water storage errors cannot be obtained by simply averaging the gridded errors for an arbitrarily shaped region. We use the squared exponential covariance function to estimate the statistical covariance between two grids as

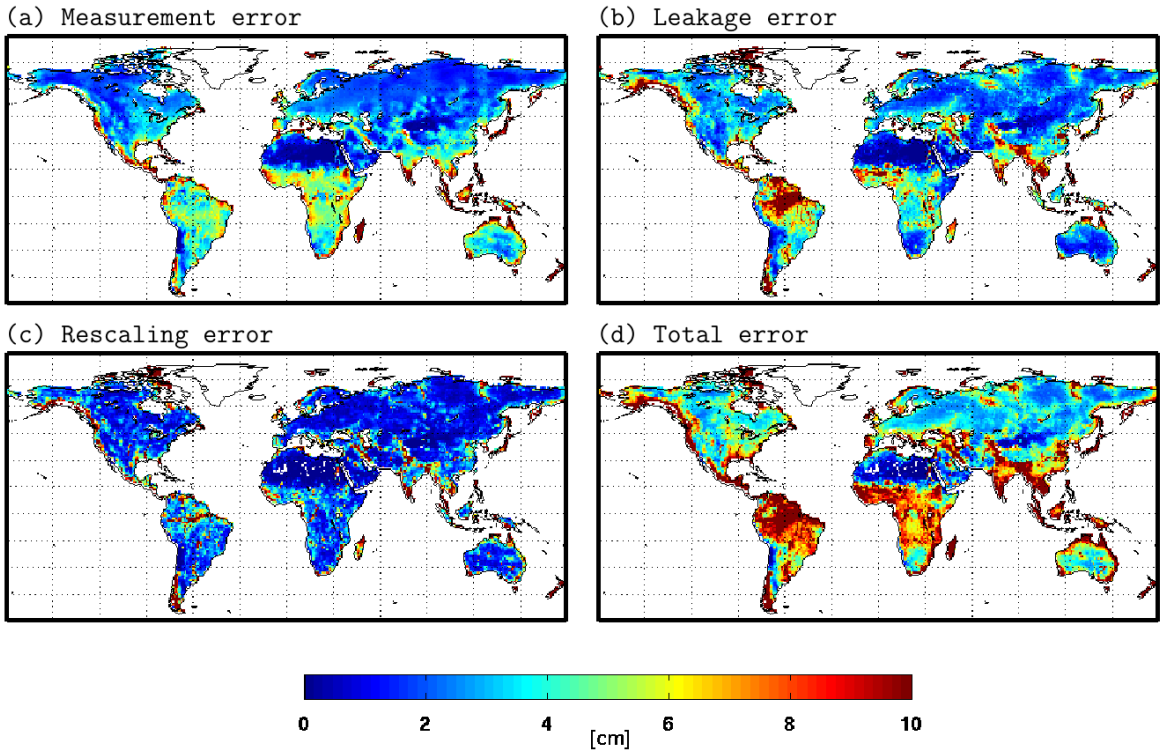

Figure 3.13: Estimates of GRACE-based TWS errors calculated at 1° grid-scale level: measurement errors (a), leakage errors (b), re-scaling errors (c), and total errors (d).

proposed by Landerer & Swenson (2012) and estimate the error variance of the regional mean TWS estimate by the following equation:

$$var = \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \theta_i \theta_j exp(\frac{-d_{ij}^2}{2d_0^2}), \tag{3.10}$$

where θ_i is the error for grid point i, w_i represents the area weight at the grid i, d_{ij} is the distance between the two points and d_0 is the parameter in the Gaussian window representing the de-correlation length scale. We choose 300 km, 100 km, 10 km as d_0 for measurement error, leakage error and re-scaling error separately by fitting the error budget from the gridded data set to the ones obtained from estimating errors directly at basin-scale level (Chen *et al.*, 2007; Klees *et al.*, 2007) as shown in Fig. 3.15. Both methods provide generally consistent results down to a level of about 20% , indicating that the errors at 1° spatial resolution might indeed serve as a starting point for the derivation of realistic errors for regions of arbitrary shape.

Since only diagonal elements of the covariance matrix have been provided for GRACE release 05a of GFZ Potsdam, the error correlations between the individual Stokes coefficients are

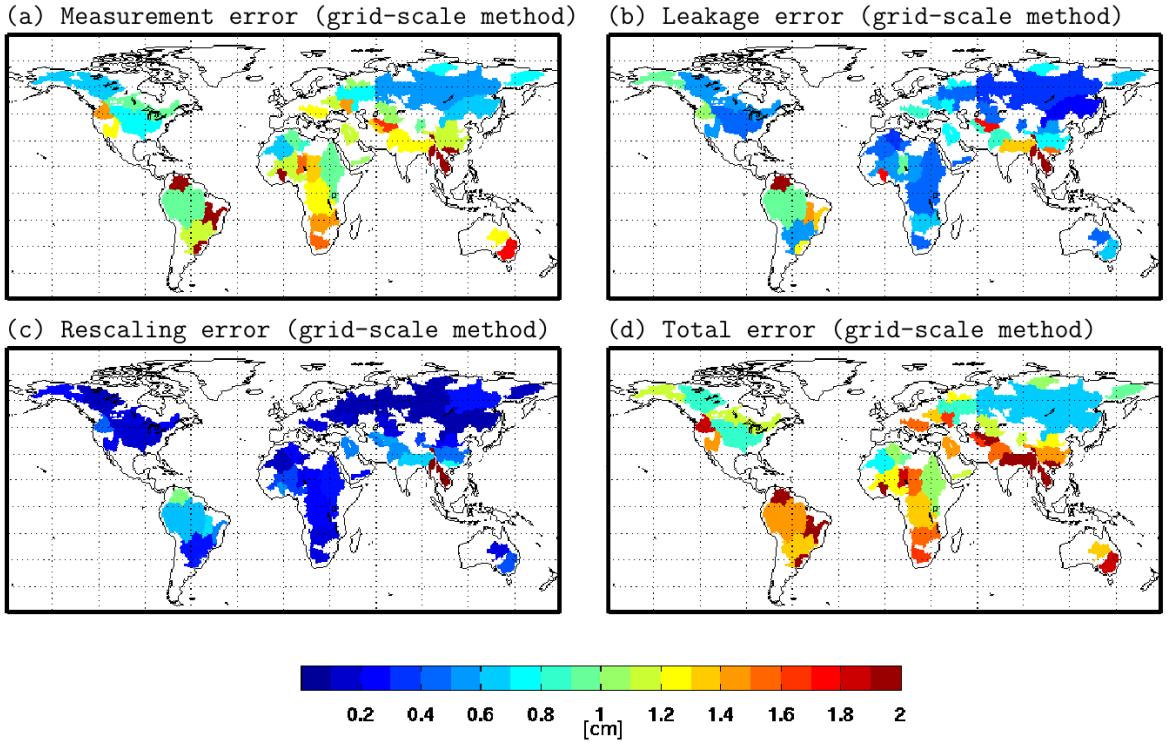

Figure 3.14: Estimates of GRACE-based TWS errors for the 50 largest discharge basins derived from 1° grid point estimates: measurement errors (a), leakage errors (b), re-scaling errors (c), and total errors (d).

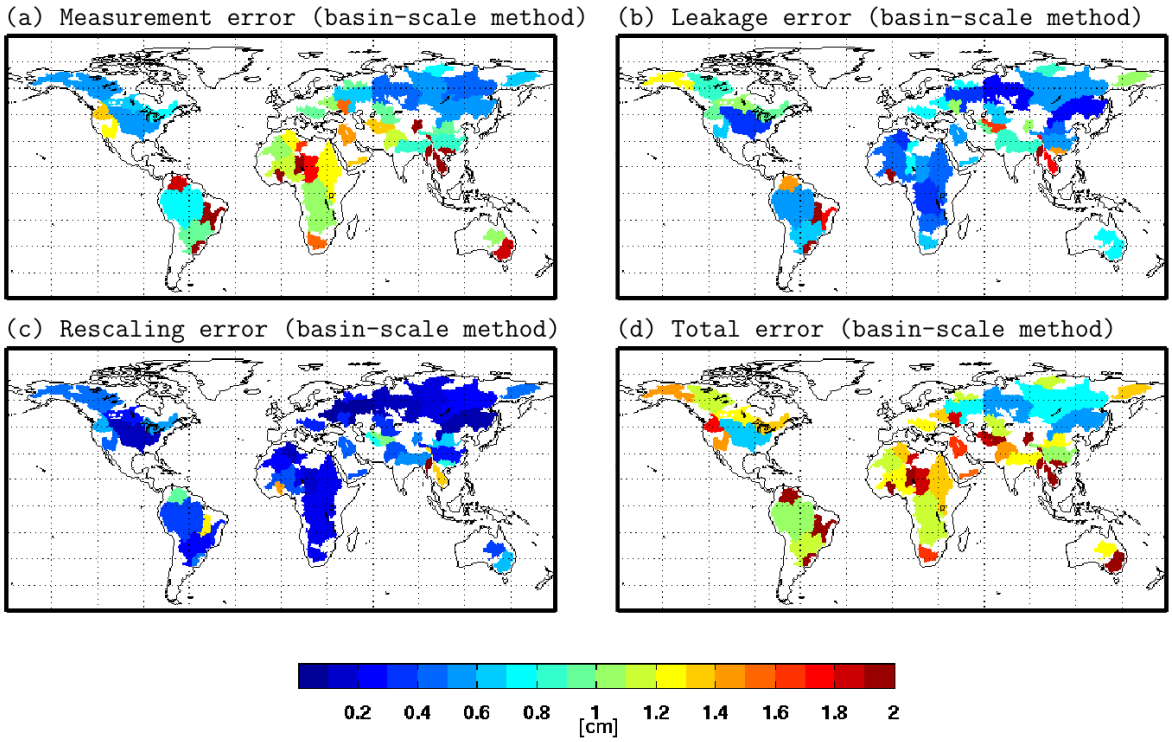

Figure 3.15: Estimates of GRACE-based TWS errors for the 50 largest discharge basins directly calculated at the basin-scale level out of the Level-2 Stokes coefficients: measurement errors (a), leakage errors (b), re-scaling errors (c), and total errors (d).

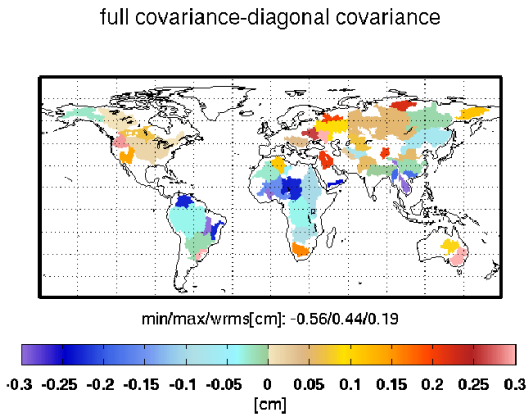

Figure 3.16: Changes in measurement errors for ITSG-Grace2014 when covariances are taken into account in addition to the usually considered variances.

typically ignored. To assess the impact of neglecting the error correlations, error estimates from both the diagonal and full variance-covariance matrices of the ITSG-Grace2014 release (Mayer-Gürr *et al.*, 2014) are calculated (Fig. 3.16). The ITSG-Grace2014 measurement errors propagated from only the diagonal part show consistent results with GFZ RL05a. The differences between the measurement errors from the diagonal and full covariance matrix reach 0.56 cm at some basins (Fig. 3.16). Generally, when the error correlations are neglected, the measurement errors at the lower latitudes are overestimated, while at higher latitudes they are underestimated. GFZ is planning to provide the error variance-covariance matrix as well (Ch. Dahle, personal communication, 2015), which we believe is necessary to further improve the reliability of the TWS error estimates.

3.3.5 Comparison with gridded TWS results from Tellus

We now compare the basin-averaged TWS time series from our calculation with gridded TWS products downloaded from the Tellus website, which are based on GRACE Stokes coefficients truncated at degree and order 60, destriped and smoothed by using a 300km Gaussian filter following (Swenson & Wahr, 2006) and then re-scaled by scaling factors derived from NCAR's CLM4.0 land surface model (Landerer & Swenson, 2012). The RMS of the differences between the TWS variations filtered with two different methods generally lie within the bounds of the GRACE error estimates (cf. Fig. 3.14d), which indicates consistency between the different filtering and smoothing methods applied (Fig. 3.17a).

By comparing the two re-scaled TWS time series, we find several basins with much larger differences in particular in South America and Southeast Asia and there are ten out of fifty

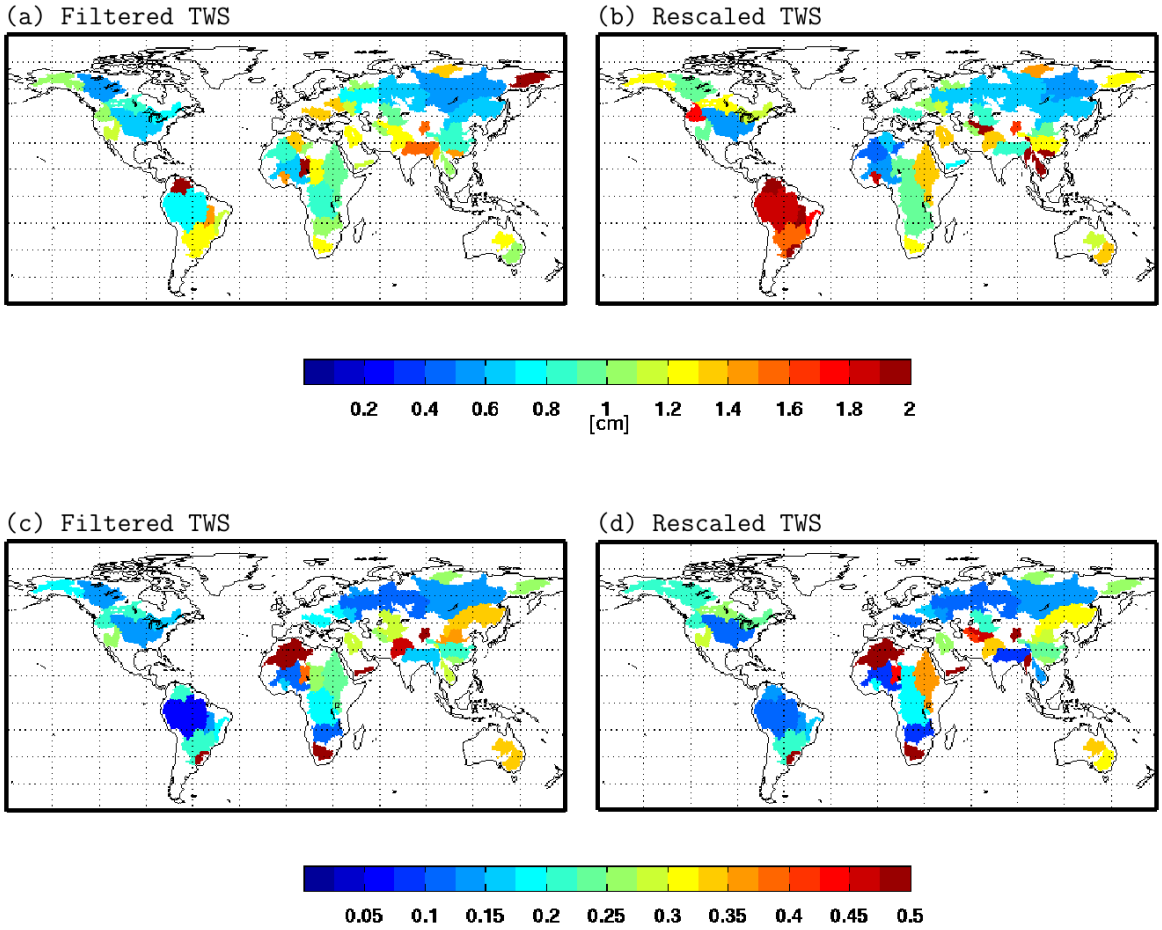

Figure 3.17: RMS of the differences between our estimated basin-averaged TWS filtered with DDK2 and the ones from Tellus for GFZ RL05a for the 50 largest discharge basins (a); and RMS of the differences between our filtered TWS re-scaled by the median scaling factors and the results from gridded TWS re-scaled by the scaling factors provided separately also on the Tellus website (b). (c) and (d) are the relative differences for (a) and (b) where the RMS of the differences are divided by the RMS of the basin-averaged TWS from Tellus.

basins where the RMS of the differences are larger than the GRACE error estimates.(Fig. 3.17b; Table 3.2). The relative differences as the percentage of the RMS of the basin-averaged TWS from Tellus are shown as well (Fig. 3.17c, d). The large relative differences mainly occur at small basins. By taking Irrawaddy as an example, we present time series of the TWS variations re-scaled by the scaling factors from all five hydrological models and the median value and compare them with the re-scaled TWS from Tellus (Fig. 3.18). In this catchment, the TWS variations re-scaled by different scaling factors from the models show a comparably large spread, but they are consistent in a way that all the scaling factors are larger than one indicating signal loss caused by filtering. Besides, the TWS time series re-scaled by median scaling factors generally lie in the middle. The results from Tellus show, however, a damping effect from the scaling factors. The TWS differences caused by applying two different scaling factors reach 10 cm, much larger than the total error (Table 3.2). This could be related to both the shape of the basin and also to the hydrological signal within and around the basin. The Irrawaddy catchment is rather elongated and shares a long border with its neighboring basins which also have high water storage variability. Both factors make it highly susceptible to spatial leakage effects and thereby vulnerable to uncertainties in simulated water storage in and outside the basin. If only a single model is used for the re-scaling, inherent uncertainties of this processing step remain inaccessible and might lead to additional errors in the GRACE-based TWS series that are not accounted for in the associated error estimates.

Table 3.2: Re-scaling factors estimated from the WGHM, LSDM, GLDAS, JSBACH, and MPI-HM models for the 50 largest discharge basins of the world. Additionally provided are the median re-scaling factors subsequently applied in the GRACE processing; the variation (VAR) coefficients of the re-scaling factors quantifying the uncertainty associated with the re-scaling; total error and signal amplitude of the basin-averaged TWS (unit: cm) and the GRACE signal-to-noise ratio; the RMS differences between rescaled TWS from Tellus and the calculation in this work with median scaling factors. Basin names are taken from the Simulated Topological Network data-set (STN-30p; Vörösmarty *et al.*, 2000b). GHAAS#Number indicate unidentified basins, where the number is the unique integer identifier adopted in STN-30p.

Discharge	re-scaling coefficients						VAR-	Total	Signal	GRACE-	Tellus-
Basin	LSDM	WGHM	JSBACH	MPI-HM	GLDAS	Median	coeff.	error	amp.	SNR	Median
Amazon	1.03	1.01	1.02	0.96	1.01	1.01	0.02	1.46	14.28	9.76	1.81
Nile	1.08	1.16	0.97	1.05	0.91	1.05	0.08	1.06	3.46	3.26	1.33
Zaire	1.08	0.96	1.04	1.01	0.99	1.01	0.04	1.32	5.05	3.82	0.98
Mississippi	1.00	1.00	0.97	0.95	0.97	0.97	0.02	0.86	5.67	6.60	0.59
Amur	1.01	0.99	0.96	0.96	0.94	0.96	0.03	0.68	2.17	3.18	0.67
Parana	1.14	1.10	1.06	1.08	1.01	1.08	0.04	1.32	5.92	4.50	1.51
Yenisei	0.96	0.92	1.03	0.98	0.97	0.97	0.03	0.68	4.52	6.67	0.64
Ob	0.94	0.98	1.02	0.96	0.99	0.98	0.03	0.68	5.64	8.31	0.66
Lena	0.94	0.88	1.00	0.98	0.96	0.96	0.04	0.68	4.12	6.01	0.60
Niger	1.04	1.01	1.03	1.05	0.99	1.03	0.02	1.29	6.37	4.93	0.54

Continued on next page

Table 3.2 – *Continued from previous page*

Discharge	re-scaling coefficients						VAR-	Total	Signal	GRACE-	Tellus-
Basin	LSDM	WGHM	JSBACH	MPI-HM	GLDAS	Median	coeff.	error	amp.	SNR	Median
Zambezi	1.02	0.98	1.02	1.00	0.97	1.00	0.02	1.57	10.66	6.80	0.91
GHAAS #14	1.42	1.74	1.77	0.70	0.99	1.42	0.30	0.77	0.68	0.88	0.47
Chang Jiang	1.37	0.98	0.89	0.98	0.89	0.98	0.19	1.49	4.61	3.09	1.26
Mackenzie	0.69	1.02	1.00	0.97	0.93	0.97	0.13	0.83	5.15	6.20	0.92
Ganges	0.96	1.09	1.00	1.00	0.97	1.00	0.05	1.94	11.62	5.99	0.81
Chari	1.19	1.08	1.01	1.02	0.97	1.02	0.08	1.50	5.14	3.42	0.92
Volga	1.00	0.99	0.98	0.98	0.98	0.98	0.01	0.84	7.07	8.43	0.68
St. Lawrence	1.02	0.98	0.85	0.75	0.79	0.85	0.13	1.14	5.63	4.94	1.14
Indus	1.17	0.87	1.00	0.95	0.89	0.95	0.12	1.54	3.72	2.42	1.32
Syr-Darya	1.20	0.92	1.01	0.99	1.00	1.00	0.10	1.12	4.08	3.65	0.84
Nelson	0.93	1.01	0.94	0.87	0.94	0.94	0.05	1.12	4.28	3.82	1.26
Orinoco	0.99	1.09	1.07	1.08	0.93	1.07	0.07	3.14	14.88	4.74	1.97
Murray	1.31	1.04	1.07	1.13	0.91	1.07	0.12	1.88	5.13	2.73	1.38
Great Artesian	1.76	1.17	0.97	0.97	1.29	1.17	0.25	1.33	3.54	2.67	1.10
Shatt el Arab	1.19	0.90	1.15	1.01	0.88	1.01	0.12	1.49	5.67	3.81	1.40
Orange	1.00	1.02	1.20	1.08	0.91	1.02	0.09	1.65	2.75	1.67	1.29
Huang He	1.28	0.52	0.99	1.05	1.02	1.02	0.25	1.28	3.00	2.35	0.91
Yukon	1.08	0.83	0.90	0.88	1.12	0.90	0.15	1.19	9.13	7.68	1.24
GHAAS #34	1.14	1.06	1.40	0.84	0.90	1.06	0.19	1.04	1.13	1.09	0.67
Colorado (Ari)	1.39	0.82	1.04	1.11	0.97	1.04	0.18	1.41	3.93	2.78	0.99
Danube	1.01	1.04	0.96	1.00	0.95	1.00	0.03	1.50	7.45	4.96	0.86
Mekong	1.11	1.07	1.02	1.08	0.90	1.07	0.08	3.86	14.42	3.73	2.32
Tocantins	1.06	1.09	1.04	1.21	1.00	1.06	0.07	2.81	16.72	5.95	2.69
Columbia	0.98	0.99	0.98	1.07	0.92	0.98	0.05	1.85	9.87	5.32	1.76
GHAAS #49	0.80	0.97	0.94	0.84	0.92	0.92	0.08	1.86	3.06	1.65	0.97
Kolyma	0.99	0.78	1.03	0.92	1.04	0.99	0.11	0.98	4.55	4.65	1.27
Sao Francisco	1.06	1.04	0.99	1.09	1.04	1.04	0.03	2.81	9.65	3.43	1.80
Amu-Darya	1.32	1.27	1.11	1.20	0.95	1.20	0.13	2.47	7.27	2.95	2.34
Dnepr	0.99	0.98	0.97	0.99	0.98	0.98	0.01	1.32	6.37	4.81	1.09
Don	1.14	1.07	1.12	1.12	1.05	1.12	0.03	1.68	8.67	5.17	1.16
GHAAS #50	1.01	0.85	1.72	0.47	0.81	0.85	0.48	1.20	0.89	0.74	0.70
Zhu jiang	1.07	1.18	1.19	1.37	1.11	1.18	0.10	3.18	7.92	2.49	2.39
Irrawaddy	1.32	1.43	1.40	1.06	1.26	1.32	0.11	4.18	17.44	4.17	10.07
Volta	0.63	0.88	0.98	1.06	0.97	0.97	0.18	3.36	9.94	2.96	1.83
GHAAS #54	0.72	0.74	1.28	0.83	1.10	0.83	0.27	1.66	3.25	1.96	1.03
Khatanga	0.87	0.94	0.87	0.89	0.93	0.89	0.04	1.01	5.48	5.42	1.43
Dvina	1.04	1.03	1.10	1.01	1.00	1.03	0.04	1.25	7.58	6.05	1.02
Urugay	1.21	1.12	1.30	1.21	1.16	1.21	0.06	2.70	7.28	2.70	3.00
Qarqan	0.49	1.07	0.23	1.04	1.33	1.04	0.54	1.14	1.34	1.18	1.63
GHAAS #75	0.26	1.02	0.77	0.51	0.83	0.77	0.44	0.94	0.75	0.80	0.48

3.3.6 Summary and Conclusions

Globally gridded estimates of terrestrial water storage anomalies have been processed from the GRACE release 05a monthly-mean gravity fields from GFZ Potsdam (Dahle *et al.*, 2012) by applying state-of-the-art post-processing methodologies. The de-correlation filter of DDK2

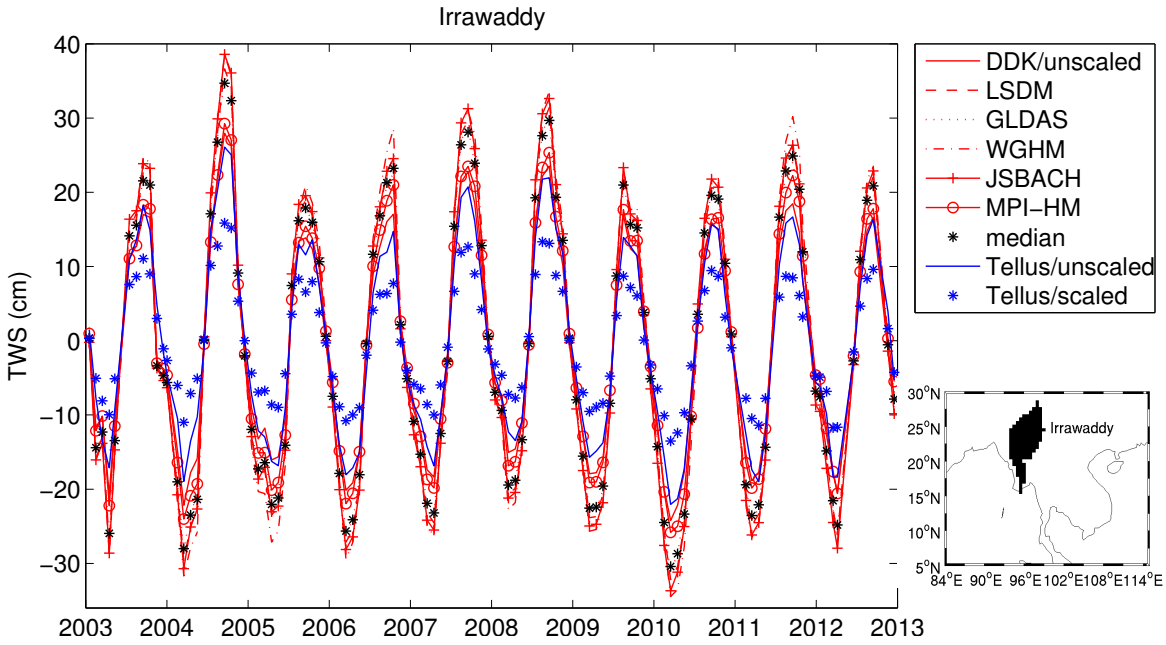

Figure 3.18: The comparison of the basin-averaged TWS time series from the calculation in this work and from Tellus for the Irrawaddy catchment. DDK/unscaled indicates the TWS variations that are filtered by DDK2; LSDM indicates the DDK2 filtered TWS re-scaled by scaling factors from LSDM, and the same for WGHM, GLDAS, JSBACH and MPI-HM; median is the TWS variations re-scaled by the median scaling factors from the five models; Tellus/unscaled is the basin average of the gridded TWS from Tellus website and Tellus/scaled is the re-scaled TWS from Tellus.

has been chosen with the goal of minimizing signal loss while maximizing noise reduction. Re-scaling factors required to account for signal loss during filtering were obtained from the median values of a small ensemble of five global models in order to make the re-scaling more robust against particular weaknesses of a single model. We therefore intend to include those globally gridded TWS data-sets as an additional Level-3 product into the ICGEM website accessible at "icgem.gfz-potsdam.de/ICGEM", so that it will be routinely updated as soon as new Level-2 gravity fields become available, and thereby contributes to a better accessibility of near real-time GRACE information to users that are not willing or not able to process Stokes coefficients themselves.

In addition to the monthly TWS estimates, we prepared realistic globally gridded error estimates by assessing individually the contributions of measurement errors, leakage errors, and re-scaling errors. The error estimates account for spatial correlations and yield largely consistent results for most basins when compared to estimates that are directly derived from the spherical harmonics representation. Thus, errors might guide users in selecting proper averaging regions and remind them that GRACE is in particular sensitive to the largest spatial scales as demonstrated by the fact that the SNR values generally decrease when going from large basins to small basins.

The RMS of the differences between the filtered TWS from our calculation and those from Tellus generally lie below the TWS error level estimated, which underlines the consistency of the two post-processing strategies. Larger differences found at certain basins between the rescaled TWS time-series, however, emphasize the importance of model-based information required to account for spatial leakage. Since global land models perform differently in simulating the TWS variability in different areas of the world, an ensemble of multiple models is helpful to make scaling factors less affected by deficiencies in certain models. In view of the important role of such models for the GRACE processing, a detailed evaluation of the quality of such models and their systematic weaknesses is recommended.

Currently, the GRACE mission has been in orbit for more than 13 years and continues to provide monthly-mean snap-shots of the global gravity field. The GRACE Follow-On mission is already in its implementation phase and scheduled for launch in 2017 (Flechtner *et al.*, 2014), thereby improving the prospects of establishing a long-term monitoring of global TWS variability with gravimetric methods. The GRACE mission has already contributed unique observations to five out of six current Grand Challenges of the World Climate Research Programme: (1) Melting Ice and Global Consequences (Sasgen *et al.*, 2010); (2) Climate Extremes (Reager *et al.*, 2014); (3) Regional Sea-Level Change (Chambers *et al.*, 2010); (4) Water Availability (Famiglietti & Rodell, 2013); and also (5) Decadal Climate Prediction

(Zhang *et al.*, 2015). The observing concept is therefore in a good position to be considered as a contribution to the 'Essential Climate Variables' (Hollmann *et al.*, 2013) as defined by the World Meteorological Organization. To foster more applications of satellite gravimetry in scientific fields like hydrometeorology and climatology, conveniently pre-processed data-sets as described in this thesis are an essential prerequisite.

Chapter 4

Validation of hydrological models

Published as: Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., and Thomas, M. (2016). Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, Hydrol. Earth Syst. Sci., 21, 821-837, doi:10.5194/hess-21-821-2017.

4.1 Hydrological model simulations

Hydrology is the study of the movement, distribution, and quality of water on Earth. It encompasses both the hydrological cycle and water resources and is an important indicator of global change. Besides, the observations of the water storage processes will also help to provide reliable predictions for future water availability and hazardous risks. Hydrological models are simplified, conceptual representations of the continental water cycle and its components. They also help to understand how water transports on the continents. To simulate the water cycle, some factors have to be applied to constrain the hydrological models, such as precipitation and its transfer to other Earth's subsystems as atmosphere and oceans by the processes of evaporation and runoff. Besides, other water characteristics like soil properties, vegetation cover, watershed topography, soil moisture content and groundwater are also considered. Nowadays, the total water storage (TWS) variations become an important variable in monitoring the stability and dynamic behavior of the water cycle and in evaluating large-scale models (Güntner *et al.*, 2007; Werth, 2010).

The terrestrial water balance describing the partitioning of precipitation (P) into evapotranspiration (E) and runoff (R) is commonly expressed as:

$$TWSC = TWS_N - TWS_{N-1}, \tag{4.1}$$

and

$$TWSC = P - E - R \tag{4.2}$$

where the term TWSC represents the change in water storage which is mainly from the sum of the individual components:

$$TWS = GW + SM + SW + SWE, \tag{4.3}$$

such as groundwater (GW), soil moisture (SM), surface water (SW) and snow water equivalent (SWE).

A selection of five different models is used in this thesis, ranging from complex to conceptional ones (Table 4.1). Thus, the range of uncertainty in the simulation of water storage can be represented more reliably than using just a single model. In order to ensure that this spread between the simulations is indeed related to the representation of physics in the model, four of the models are forced with the WATCH Forcing Data ERA Interim (WFDEI) data-set developed during the WATCH project (Weedon *et al.*, 2011) applied to ERA-Interim re-analysis data (Dee *et al.*, 2011). This WFDEI meteorological forcing dataset is a quasi-observation which combines the daily variability of the ERA-Interim re-analysis with monthly in situ observations such as temperature and precipitation (Weedon *et al.*, 2014). There are two precipitation products available from WFDEI: corrected by using (1) the Climate Research Unit at the University of East Anglia (CRU) observations; and (2) the Global Precipitation Climatology Centre (GPCC) observations. Since the WFDEI data sets incorporating the CRU-corrected precipitation products cover a longer time span, they are used in our thesis and referred to as WFDEI-CRU.

4.1.1 WGHM

The WaterGAP Global Hydrological Model (WGHM) is part of the Water-Global Assessment and Prognosis model (WaterGAP; Döll *et al.*, 2003). WGHM is a water-balance model with conceptual formulations which simplify the most important hydrological processes on large scales. It is optimized by tuning runoff coefficients against observed river discharge in a station-based calibration (Hunger & Döll, 2008). The model simulates the continental water cycle including water storage compartments such as soil moisture within the effective root zone of vegetated areas, groundwater canopy water, snow and surface water in rivers, lakes, reservoirs and wetlands. The simulations are also supplied by cell-based information on the properties of soil, land cover, and locations of reservoirs, lakes, and, wetlands. Human water consumption is also considered in the water cycle. The potential evapotranspiration

is based on Priestley and Taylor's approach which requires net radiation as input data. The equation is adjusted by a coefficient that differentiates between humid (average relative humidity of 60% or more) and arid regions (average relative humidity less than 60%). The actual evapotranspiration is calibrated by the open water albedo and sublimation of snow. The effect of snow is simulated by a degree-day approach, where the degree-day factor is based on the type of land cover and is calibrated by a multiplicative factor. Additional effects on snow storage processes are adjusted by a cell-averaged snow freeze and melting temperature. The latest version of WGHM as calibrated for WFDEI-GPCC forcing (version 2.2 STANDARD; Müller Schmied *et al.*, 2014) has been applied. In our thesis, however, we use WFDEI-CRU forcing without re-calibrating the model due to the computational issues.

4.1.2 LSDM

The land surface discharge model (LSDM; Dill, 2008) is based on the Simplified Land Surface Scheme (SL-Scheme) and the Hydrological Discharge Model (HD-Model; Hagemann & Gates, 2003, 2001) from the Max-Planck-Institute for Meteorology. The SL-Scheme includes a soil bucket scheme for the computation of the vertical water balance and uses a simple snow scheme based on the degree day approach. The daily fields of runoff and drainage from the SL-Scheme are passed down to the HD-Model, which is a state of the art river routing model. The code has been tailored to enable the simulation of globally the vertical and lateral water transport and storage on the continental land surface for geodetic applications, that include the derivation of Effective Angular Momentum Functions of the continental hydrosphere to interpret and predict changes both in the Earth rotation (Dobslaw *et al.*, 2010; Dill & Dobslaw, 2010); and of vertical crustal deformations as observed from GPS permanent stations (Dill & Dobslaw, 2013). The global water storage variations contain water in rivers, lakes, wetlands, groundwater and soil moisture, as well as water stored in snow and ice.

4.1.3 JSBACH

JSBACH (Raddatz *et al.*, 2007; Brovkin *et al.*, 2009) is a land surface model and forms together with ECHAM6 (Stevens *et al.*, 2013) and MPIOM (Jungclaus *et al.*, 2013) the current Max-Planck-Institute for Meteorology's Earth System Model (MPI-ESM). As part of the MPI-ESM, JSBACH includes interactive vegetation and a 5-layer soil hydrology scheme to provide the lower atmospheric boundary conditions over land, particularly the fluxes of energy, water and momentum. Snow is treated as external layers above the soil column.

With increasing snow depth in winter, new layers are added up to maximum of five snow layers. The top four layers are always 5 cm in thickness, while the bottom layer is unlimited in size. Soil moisture in deep layers below the root zone is simulated and buffers extreme soil moisture conditions in the layers above. For this thesis, however, JSBACH was used in an offline mode without interactive coupling to the other MPI-ESM compartments, but driven by prescribed WFDEI-CRU atmospheric forcing.

4.1.4 MPI-HM

Finally, the Max Planck Institute of Meteorology's Hydrology Model (MPI-HM; Stacke & Hagemann, 2012) is a global hydrological model. Similar to LSDM, it is also adapted from two sub-components, the SL-Scheme and HD-Model. Its water flux computations are of similar complexity to land surface models, but it does not account for any energy fluxes. Additionally to precipitation and temperature, it requires potential evapotranspiration as input which also was derived from the WFDEI using the Penman–Montheith equation similar to the Weedon *et al.* (2011) study. TWS from MPI-HM is simulated as the sum of soil moisture in the root zone, snow and surface water.

4.1.5 GLDAS

The Global Land Data Assimilation System (GLDAS) operated at the National Center for Environmental Prediction (NCEP) (Rodell *et al.*, 2004) is a land surface simulation system which incorporates ground and space-based observations of the global water and energy cycle into the land surface models. GLDAS is driven by the meteorological forcing data come from the Global Data Assimilation System (GDAS) (Derber *et al.*, 1991), supplemented with a down-scaled version of the NOAA Climate Prediction Center's Merged Analysis of Precipitation (CMAP) (Xie & Arkin, 1997) and downward radiation from the Air Force Weather Agency (Kopp & Kiess, 1996). We analyze a single realization of GLDAS that is based on the Noah land surface model. The continental water storage in GLDAS includes the soil moisture storage in four layers within the first 2 m of the ground, snow water equivalent and canopy interception storage, while groundwater and surface water are not accounted for.

Some of the main characteristics of the numerical models are presented in Table 4.1, which provide more information on how models are different with each other. For instance, although soil moisture and snow water are included in all models, surface water and groundwater are simulated differently. JSBACH is the only model which does not include surface water. Groundwater is simulated by WGHM, where the anthropogenic impact such as groundwater

Table 4.1: Overview of the main characteristics of the five numerical models particularly considered in this thesis.

Model name	Meteorological forcing variables	Storage compartments included	Soil moisture depth	Snow	Potential Evapotranspiration
LSDM	Precipitation, temperature	subsurface water (root zone), snow, surface water	bucket scheme without a depth	degree day	Thornthwaite
WGHM	Precipitation, temperature, shortwave radiation	subsurface water (root zone+ groundwater), snow, surface water	varies with rooting depth of land cover	degree day	Priestley–Taylor
JSBACH	Precipitation, temperature, wind, shortwave and longwave radiation, surface air	subsurface water (root zone+ deep layer), snow	down bedrock but at most 10 m	energy balance	physical parametrization
MPI-HM	Precipitation, wind, temperature, radiation, humidity	subsurface water (root zone), snow, surface water	bucket scheme without a depth	degree day	Penman–Montheith
GLDAS	GDAS meteorological data, CMAP precipitation	canopy, soil moisture, snow	Within first 2 m of the ground	-	-

abstraction is also considered. JSBACH does not include groundwater explicitly. However, soil moisture in deep layers below the root zone is simulated and buffers extreme soil moisture conditions in the layers above. Thus, some of the characteristics of real groundwater are considered. We use the term subsurface water for both soil moisture and groundwater.

LSDM, WGHM, MPI-HM are provided on a 0.5° by 0.5° grid, while GLDAS is on 1° by 1° resolution and JSBACH has a coarse resolution, with 1.875° spacing in longitude and irregular spacing in latitude. The mean values and the linear trends estimated over the period Jan 2003 to Dec 2012 – i.e., the common period of GRACE observations and model results – are first removed for each grid cell. Then the TWS variations are averaged over the interested basins to obtain the basin-scale TWS. For the model validation, we mainly focus on four of the models (WGHM, LSDM, MPI-HM, and JSBACH) which are forced with the same meteorological data. Since ice dynamics and glacier mass balance are not included in the numerical models applied in this thesis, water mass variations in Antarctic and Greenland are not considered throughout the reminder of this paper.

4.2 TWS Estimates from GRACE

The GRACE US–German twin satellite mission provides estimates of month-to-month changes in the gravitational field of the Earth mainly based on precise K-band microwave measurements of the distance between two low-flying satellites (Wahr, 2009) since April 2002. After correcting for short-term variability due to tides in the atmosphere (Biancale & Bode, 2006),

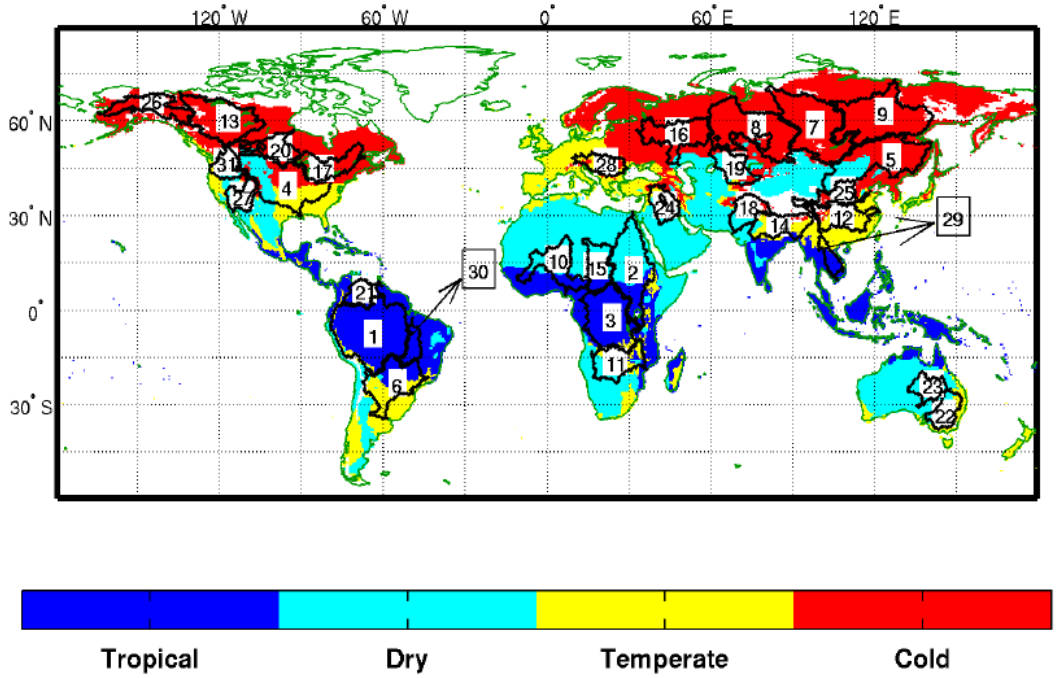

Figure 4.1: Locations of 31 globally distributed basins from the simulated topological networks (STN-30p) with underlying Köppen–Geiger climate zones. Basin IDs and names are indicated in Table 4.2.

solid Earth (Petit & Luzum, 2010) and oceans (Savcenko & Bosch, 2012), as well as due to non-tidal variability in the atmosphere and oceans (Dobslaw *et al.*, 2013) from the observations, the resulting gravity changes mainly represent mass transport phenomena in the Earth system, which are – apart from long-term trends – almost exclusively related to the global water cycle.

We use the monthly GRACE release 05a Level-2 products from GFZ Potsdam (Dahle *et al.*, 2012), which can be downloaded from the website of the International Centre for Global Earth Models (`http://icgem.gfz-potsdam.de/ICGEM`). The GRACE products are expressed in terms of fully normalized spherical harmonic (SH) coefficients up to degree and order 90, approximately corresponding to a global resolution of 2° in latitude and longitude. We apply the same post-processing steps to the GRACE data as described by Zhang *et al.* (2016). The degree-1 coefficients are added following the method of Bergmann-Wolf *et al.* (2014b). The non-isotropic filter DDK2 corresponding to an isotropic Gaussian filter with 680 km full width half maximum (Kusche, 2007; Kusche *et al.*, 2009) is applied to remove correlated errors at particular higher degrees of the spherical harmonic expansion. In order to account for signal attenuation and leakage caused by smoothing and filtering, local re-scaling factors are introduced for each grid cell. We use median re-scaling factors obtained from a small

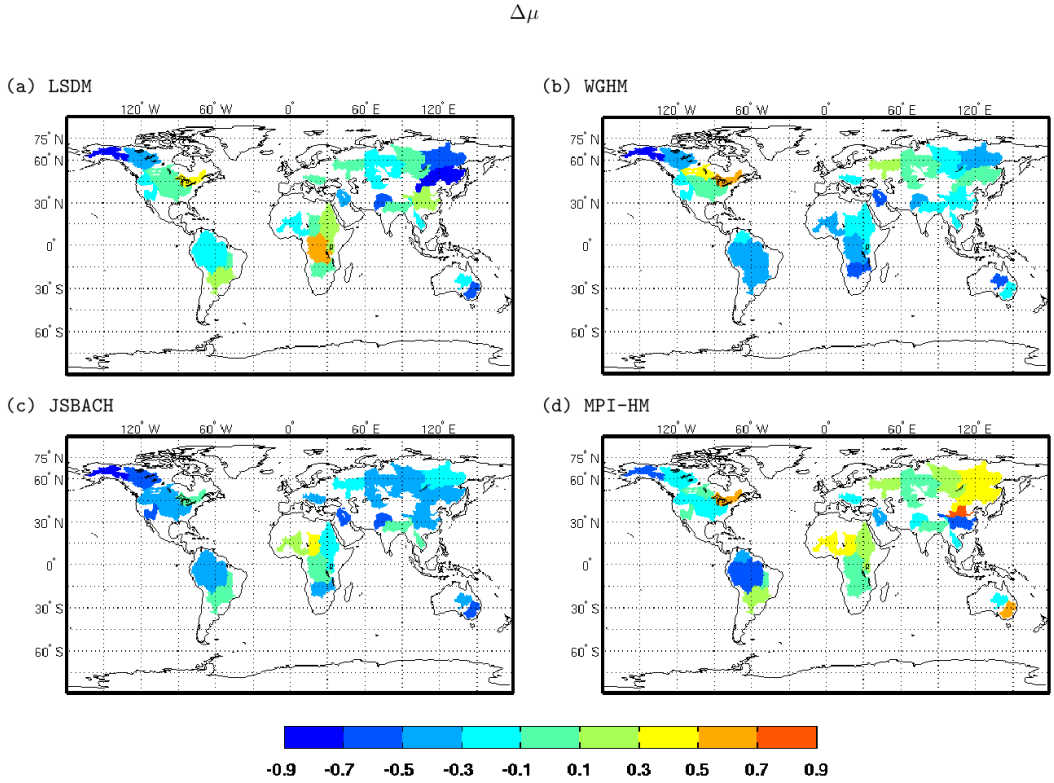

Figure 4.2: Relative amplitude differences of four hydrological model realizations with GRACE-based TWS observations.

ensemble of global hydrological models. The gridded TWS anomalies are then estimated which can be averaged over arbitrary basins. As for the model data, the linear trend is removed over the period January 2003 to December 2012. Error estimates as a quadrature of measurement error, leakage error and re-scaling error are also provided to assess the signal-to-noise ratio (SNR) of GRACE for particular basins (full details are given in Zhang *et al.*, 2016). In the case of a small signal-to-noise ratio, discrepancies between TWS from GRACE and models might also be attributed to comparatively large GRACE TWS errors.

4.3 Evaluation of TWS from model realizations with GRACE

We compare the basin-averaged TWS from GRACE with the results of four different numerical model realizations introduced above. In total, 31 globally distributed basins where the GRACE SNR is larger than 2 (see Fig. 4.1 and Table 4.1) are selected for further study. We first focus on the global statistical performance of the models compared to GRACE. For

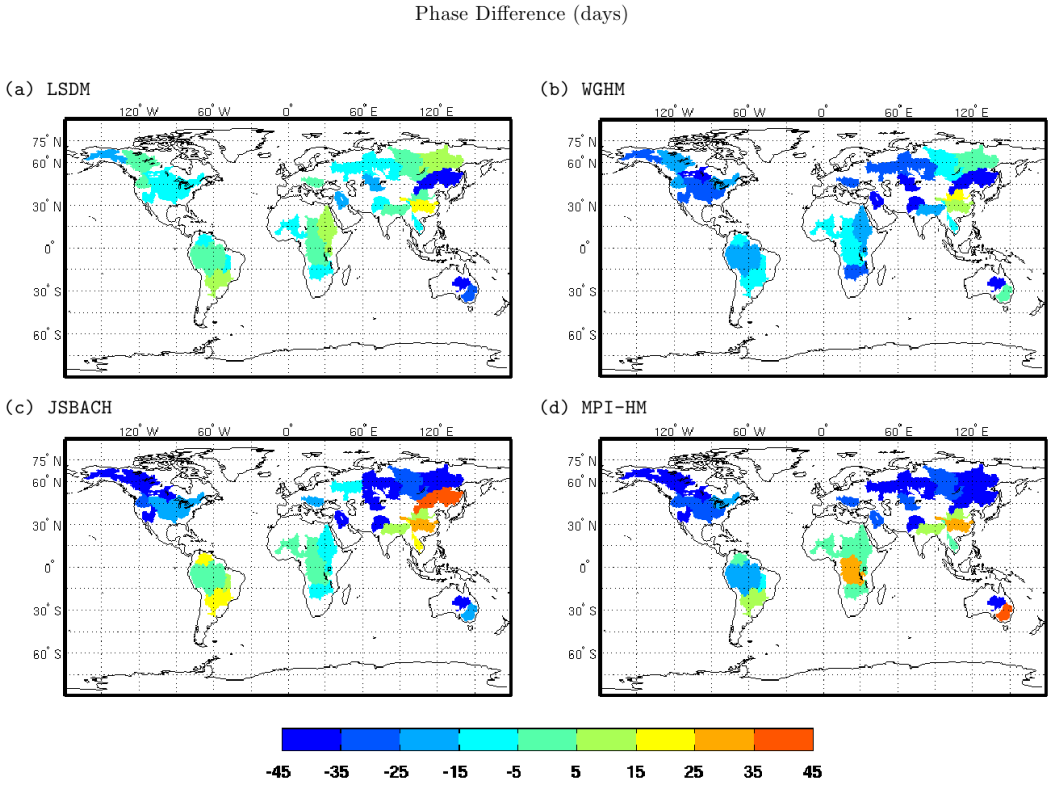

Figure 4.3: Phase differences for the annual signal of four hydrological model realizations with GRACE-based TWS observations.

these basins, evaluation metrics as suggested by Gudmundsson *et al.* (2012) that focus both on seasonal signals and year-to-year variability are applied.

4.3.1 Evaluation metrics

First, relative annual amplitude differences are calculated according to

$$\Delta\mu = (\mu_{\mathrm{M}} - \mu_{\mathrm{O}})/\mu_{\mathrm{O}}, \tag{4.4}$$

where μ_{O} is the annual amplitude of the time series of TWS variations from GRACE, and μ_{M} the annual TWS amplitudes from the different model realizations (Fig. 4.2). Second, the timing of the annual cycle is assessed using phase differences of the annual harmonic for models and observations according to

$$\Delta\phi = \phi_{\mathrm{M}} - \phi_{\mathrm{O}}. \tag{4.5}$$

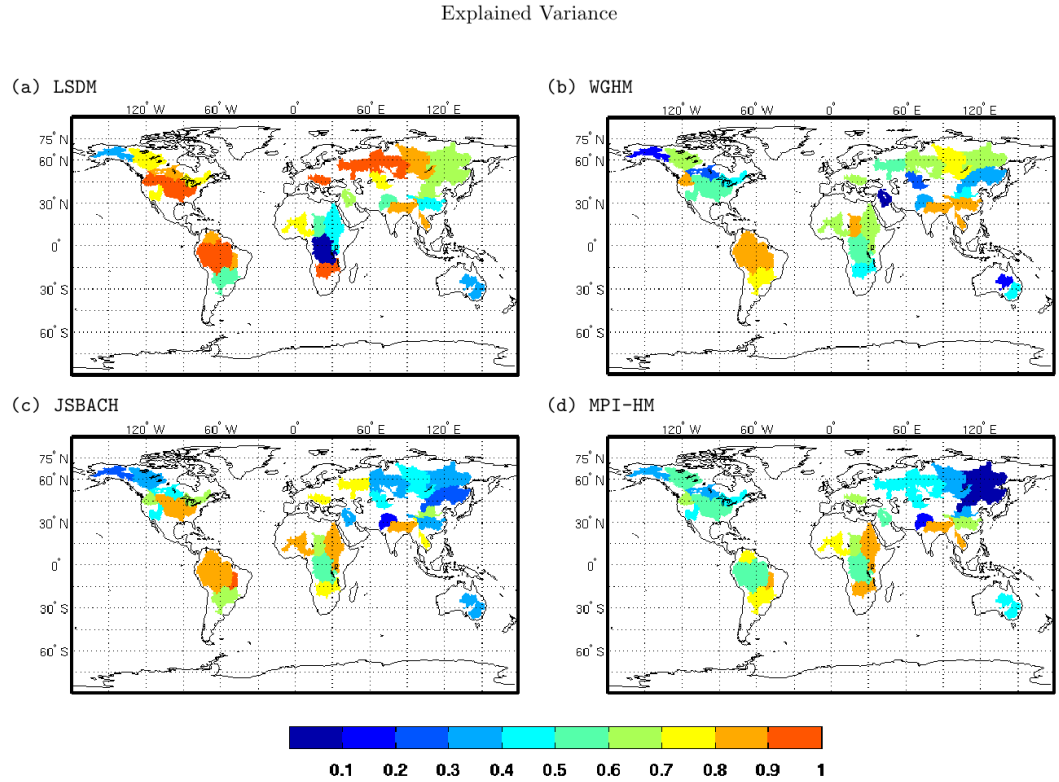

Figure 4.4: Variance of GRACE-based TWS observations that is explained by TWS as simulated in four hydrological model realizations.

If the value of $\Delta\phi$ is negative, it implies that the seasonal maximum is earlier in the year in the model than in GRACE (Fig. 4.3). Annual amplitude and phase are calculated by least square regression as follows:

$$\begin{aligned} \text{MIN} \stackrel{!}{=} (\Delta\text{TWS}(t) - (A\sin(2\pi t/T + \phi))^T (\Delta\text{TWS}(t) \\ - (A\sin(2\pi t/T + \phi)), \end{aligned} \tag{4.6}$$

where ΔTWS is the TWS anomaly time series and T is the period of 1 year. Third, the explained variances for all the model realizations are calculated:

$$R^2 = (\text{var}(\text{TWS}_\text{O}) - \text{var}(\text{TWS}_\text{O} - \text{TWS}_\text{M}))/\text{var}(\text{TWS}_\text{O}), \tag{4.7}$$

where var denotes the variance operator. Fourth, we repeat the calculation of the explained variances for TWS time series from GRACE and the models with the mean seasonal variability removed.

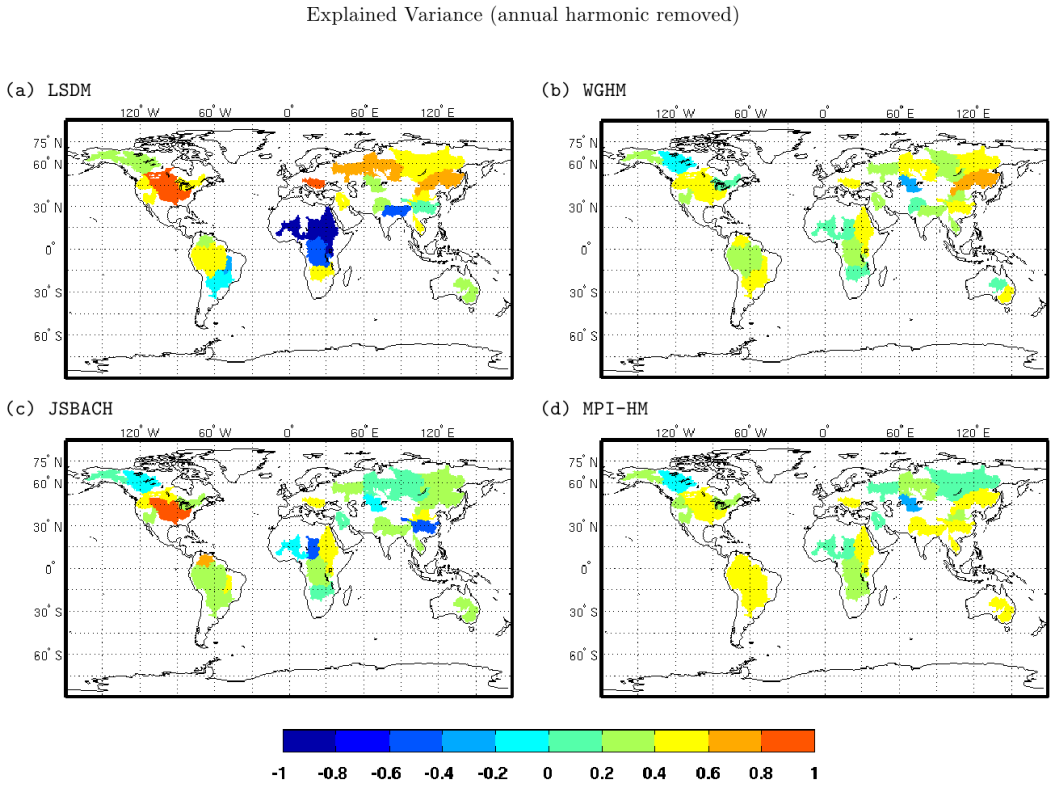

Figure 4.5: Variance of GRACE-based TWS observations that is explained by TWS as simulated in four hydrological model realizations. For both observations and model results, the annual harmonic signal has been removed.

4.3.2 Global evaluation

As shown in Fig. 4.2, the values of $\Delta\mu$ for WGHM and JSBACH are mostly negative. For JSBACH, these negative values mainly occur at mid to high latitudes of the Northern Hemisphere. WGHM underestimates the annual amplitude, especially at the low latitudes. Contrarily, MPI-HM has more basins with positive $\Delta\mu$. For LSDM, most $\Delta\mu$ values lie between -0.3 and 0.3, indicating on average better agreement of annual amplitude with GRACE. The phase difference varies more among the different models, but in most cases an earlier seasonal storage maximum is shown for the model runs relative to GRACE. There are more basins with phase difference values near zero for LSDM, while WGHM, JSBACH and MPI-HM show large differences with respect to the GRACE result, especially at high latitudes of the Northern Hemisphere (Fig. 4.3). LSDM explains the GRACE TWS variations relatively better than the other models at most basins (Fig. 4.4). Only in the Yukon, Nile, Zaire, Yangtze, Indus and the two basins in Australia are explained variances less than 50 %. Low values of explained variance also occur at the mid-latitude of the Northern Hemisphere

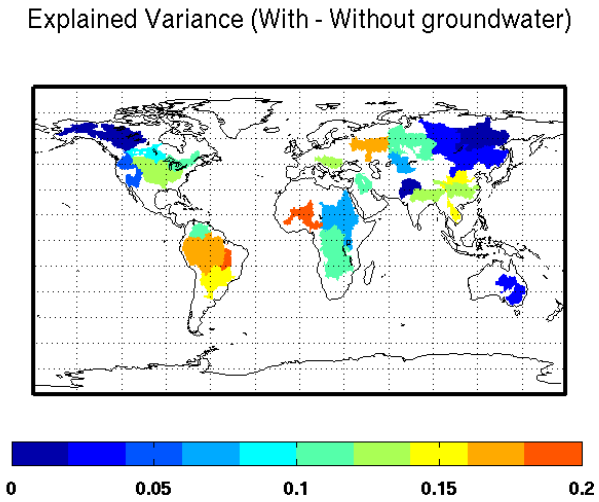

Figure 4.6: The differences between the explained variance values from WGHM with and without groundwater.

for WGHM. JSBACH and MPI-HM perform generally better at basins in Africa, but have worse results in Siberia. When the annual signal is removed, the explained variances for TWS time series from GRACE and the models are generally less than 60 % (Fig. 4.5), indicating the models's poor ability to capture the inter-annual variations. LSDM shows especially low explained variance values for many basins in Africa.

The impact from consideration of groundwater to TWS variations in WGHM is investigated by showing the differences of explained variances with and without groundwater (Fig. 4.6). The positive values indicate that WGHM with groundwater exhibits better agreement with GRACE than the one without. The large impact is mainly located at basins such as Tocantins, Niger, Huang He, Mekong and Mississippi. Only in three basins (Lena, Indus and Yukon) is the effect of groundwater consideration on the model negative.

As each metric usually focuses only on one specific property of statistical performance and has its own limitations, the time series of TWS are given for some basins with the largest deviation between GRACE and the model. We show the Yukon basin, where both WGHM and JSBACH exhibit the largest deviation of annual amplitudes from GRACE. Although the annual amplitude is simulated better by LSDM and MPI-HM, apparent negative phase differences are shown. The Amur basin is also shown, as LSDM, WGHM and MPI-HM all have the largest negative phase differences with GRACE here. Models generally capture the inter-annual signals but perform quite differently among each other and with GRACE in terms of seasonality. Almost opposite phase differences are found for these models. The smallest explained variance for MPI-HM happens at the St. Lawrence basin, where a much

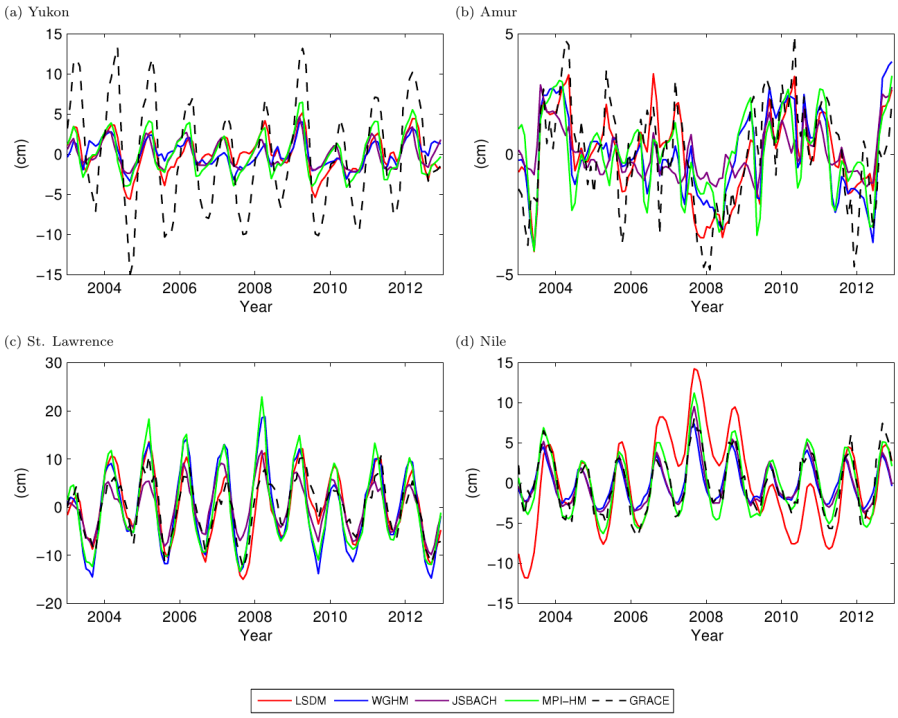

Figure 4.7: Examples of monthly TWS time series from GRACE and models for the basins with the largest deviation between model and GRACE in each of the four metrics: Relative amplitude differences (Yukon), phase differences (Amur), explained variance (St. Lawrence) and explained variance with annual harmonic signal removed (Nile).

larger amplitude and a negative phase difference compared with GRACE are found. When the annual signal is removed, models perform differently in terms of the explained variance. In the Nile basin, large inter-annual variations simulated by LSDM even lead to negative explained variance compared with the other models.

Fig. 4.8 summarizes the overall performance of each statistical metric for all the basins considered by means of box plots. The median $\Delta\mu$ for MPI-HM is almost zero where the other three values are all negative, indicating an underestimation of the annual amplitude of TWS from LSDM, WGHM and JSBACH. As shown in Fig. 4.2d, MPI-HM overestimates the TWS variations at many basins, which compensate with those underestimated values and lead to a median value at almost zero. All the models have a median phase difference below zero, with LSDM having the smallest bias and range, and MPI-HM the largest bias. This means that the TWS peaks of the models tend to proceed GRACE peaks. For the explained variance, LSDM shows the best median value, followed by WGHM, JSBACH and MPI-HM. However, when the annual signal is removed, many outliers appear in LSDM for the explained variances, while WGHM and MPI-HM show slightly better performances.

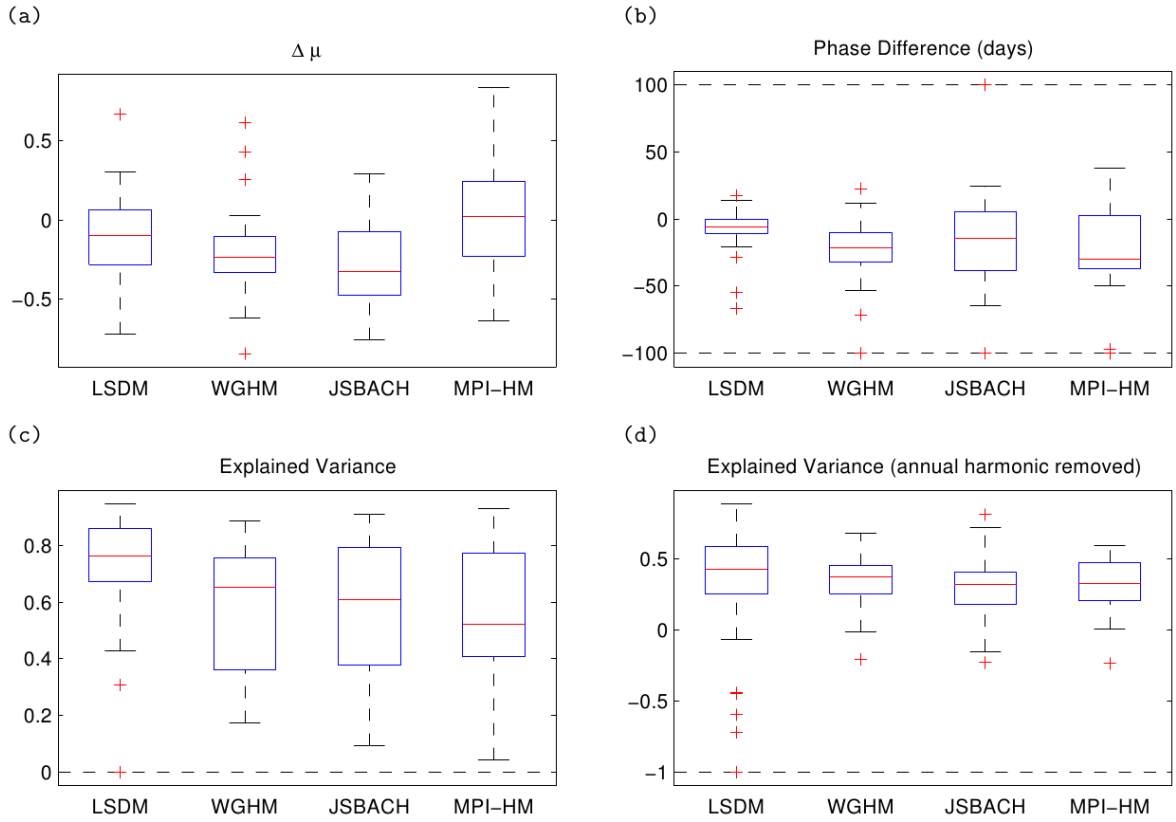

Figure 4.8: Box plots illustrating the $\Delta\mu$ (a), phase differences (b), explained variance (c) and explained variance with the annual harmonic signal removed (d) for the TWS from GRACE and models. The red horizontal line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually and set within the extreme data limits as indicated by the dashed line.

We also present the basin-averaged TWS errors from GRACE and the root mean square (rms) differences between TWS variations from GRACE and from the hydrological model runs (Table 4.2), where the largest and smallest differences are shown in bold and underlined separately. The basins are grouped according to the Köppen climate zones (Kottek & Grieser), which include tropical climates, dry climates, temperate climates and cold climates (see Fig. 4.1). For most of the basins, the GRACE errors are much smaller than the rms differences, which indicates that the main contributions to the differences arise from model uncertainties. Out of the five basins in the tropical zone, three basins have the largest differences between TWS variations from GRACE and models in LSDM. In contrast, WGHM has no largest differences in this climate zone. The smallest value, however, seems to occur randomly among the models. In the dry zone, most basins have low SNR values and the smallest rms of the TWS differences is sometimes quite close to the GRACE TWS errors. For instance, at basins

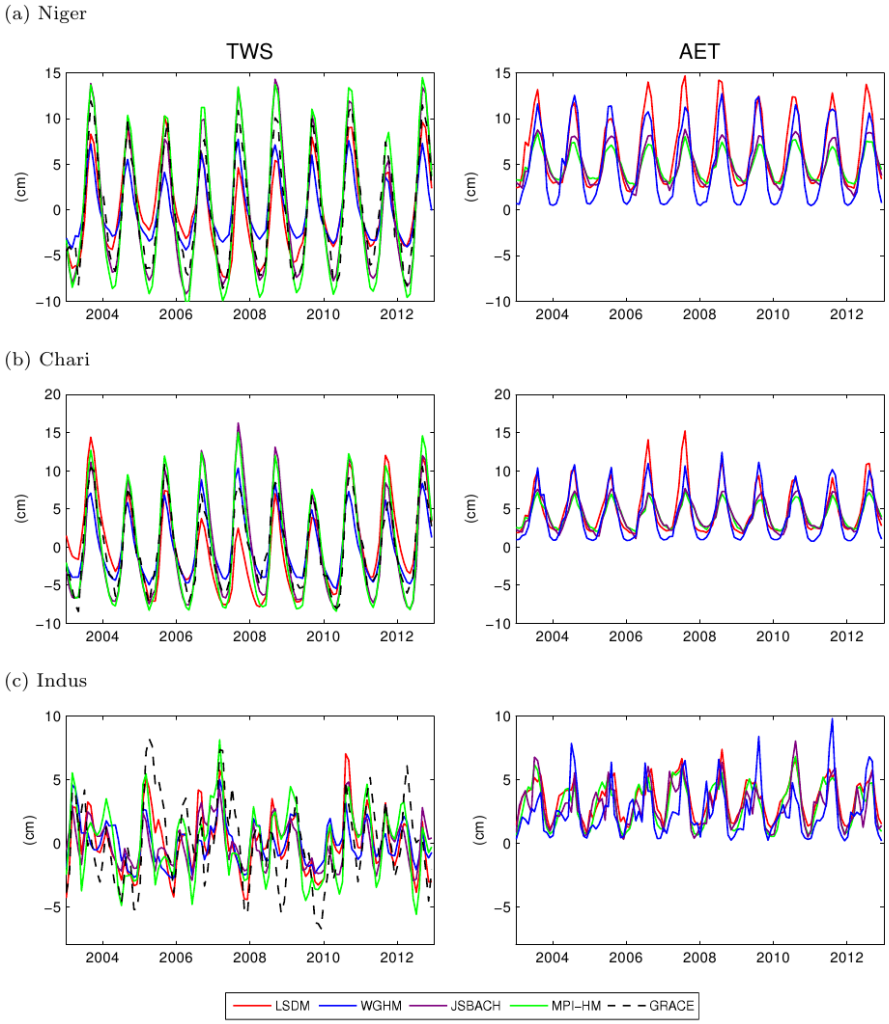

Figure 4.9: Time series of TWS (left) from GRACE and models and model simulated AET time series (right); each for three different catchments in dry zone: Niger, Chari and Indus.

like the Nile, Indus, and the two Australian basins, the GRACE SNR estimates are all below 3. Thus, it is likely that the large uncertainty in GRACE TWS estimates contributes largely to the bad agreement in these basins. Still, MPI-HM and LSDM perform comparably better, showing a smaller number of largest differences and comparably more smallest differences. In the temperate zone, WGHM has the most largest differences, while MPI-HM has the least. There is, however, no regular pattern of where the smallest difference occurs. In the cold zone, all the smallest differences happen in LSDM, whereas the largest differences mainly occur at MPI-HM and JSBACH.

The performance of the models varies from basin to basin, even within the same climate zone, which could be due to the model structure, parametrization, and also the different

Table 4.2: Characteristics of the basins shown in Fig 4.1. Bold and underlined numbers are the largest and smallest RMS differences between GRACE and models separately.

Climate Zones	Basin ID	Name	Area ($1000\,km^2$)	RMSE(cm) between TWS from GRACE and LSDM	WGHM	JSBACH	MPI-HM	GRACE TWS error(cm)	SNR
Tropical	1	Amazon	5853	4.39	6.08	5.60	**9.53**	1.46	9.76
	3	Zaire	3699	**5.26**	3.36	3.08	3.49	1.32	3.82
	21	Orinoco	1039	**6.37**	4.96	6.21	5.79	3.14	4.74
	29	Mekong	774	5.87	5.60	**6.28**	4.51	3.86	3.73
	30	Tocantins	769	**7.69**	7.49	4.99	5.45	2.81	5.95
Dry	2	Nile	3826	**4.02**	1.85	1.61	1.39	1.06	3.26
	10	Niger	2240	2.53	**2.97**	1.87	2.23	1.29	4.93
	15	Chari	1571	**2.94**	1.96	2.40	2.50	1.50	3.42
	18	Indus	1143	2.17	2.61	**3.08**	3.04	1.54	2.42
	19	Syr-Darya	1070	2.00	**3.30**	3.07	2.89	1.12	3.65
	22	Murray	1031	3.45	3.61	**3.68**	3.38	1.88	2.73
	23	Great Artesian	977	2.44	**2.67**	2.36	2.22	1.33	2.67
	24	Shatt el Arab	967	2.28	3.64	**3.67**	2.85	1.49	3.81
	25	Huang He	894	1.52	2.09	1.74	**2.38**	1.28	2.35
	27	Colorado(Ari)	807	1.90	2.59	**2.98**	2.91	1.41	2.78
Temperate	4	Mississippi	3203	1.68	**3.54**	2.36	3.45	0.86	6.60
	6	Parana	2661	**4.17**	3.03	3.59	2.81	1.32	4.50
	11	Zambezi	1989	2.89	**7.05**	4.83	3.30	1.57	6.80
	12	Chang Jiang	1794	2.58	2.05	**3.24**	3.12	1.49	3.09
	14	Ganges	1628	4.04	**4.43**	3.73	2.90	1.94	5.99
Cold	5	Amur	2903	1.20	1.73	1.88	**2.05**	0.68	3.18
	7	Yenisei	2582	1.89	2.34	3.44	**3.54**	0.68	6.67
	8	Ob	2570	1.50	3.20	**4.35**	4.14	0.68	8.31
	9	Lena	2418	2.33	2.40	3.40	**3.99**	0.68	6.01
	13	Mackenzie	1713	2.67	2.83	**3.95**	3.39	0.83	6.20
	16	Volga	1463	2.11	4.55	3.28	**5.22**	0.84	8.43
	17	St.Lawrence	1267	2.59	4.74	3.42	**4.88**	1.14	4.94
	20	Nelson	1047	1.67	**3.87**	3.19	3.31	1.12	3.82
	26	Yukon	852	5.06	5.72	**5.74**	5.29	1.19	7.68
	28	Danube	788	1.72	4.18	4.03	**4.27**	1.50	4.96
	31	Columbia	724	2.69	4.75	**6.09**	5.71	1.85	5.32

water storage components included in TWS. In order to find reasons for the different model performance, we focus on two specific areas that are dominated by snow and arid climates in more detail. There, we assess actual evapotranspiration (AET) and runoff which are the main components of the terrestrial water budget and subsequently look into the mean monthly time series of TWS and its individual storage components.

4.3.3 Actual evapotranspiration and runoff

As part of the terrestrial branch of the water cycle, actual evapotranspiration (AET) and runoff may explain part of the differences among the models in terms of storage variations. Although some large differences of AET are present, the effects on subsequently simulated TWS are damped. Especially in humid areas, no direct impact can be found. For arid basins, however, the impact from AET is more dominant. We choose three particularly affected basins (Niger, Chari and Indus) and show the AET time series from all models (Fig. 4.9). For these basins, the time series comparison shows that the smaller (or larger) AET in

the wet season leads to higher (or lower) seasonal amplitude of TWS. In addition, in these dry areas, LSDM generally exhibits enhanced AET due to high temperatures and extremely low humidity which then lead to smaller TWS variations. As exemplarily demonstrated for the Niger basin, the relatively larger AET from LSDM covering the time period 2007 to 2009 is just correspondent to the comparably smaller TWS variations. AET is calculated from the potential evapotranspiration (PET) as a function of the available amount of water. While starting with the same meteorological forcing data, PET is calculated differently by the models using various approaches. PET in the LSDM is calculated by the Thornthwait method, using only the daily temperature and a seasonal heat index that is based on monthly mean temperatures. In WGHM, PET is based on the Priestley–Taylor approach using net radiation, which in turn is computed as a function of incoming shortwave radiation, temperature and surface albedo. For MPI-HM, PET is computed in a pre-processing step based on Penman–Montheith using radiation, temperature, wind and humidity. JSBACH computes evaporation based on the energy balance by internally computing atmospheric water demand.

Fig. 4.10 displays time series comparison of runoff from the models for three basins in the tropical zone (Amazon, Orinoco and Mekong). The runoff is calculated from the models following the equation

$$R(t) = P(t) - \mathrm{ET}(t) - \mathrm{TWSC}(t), \tag{4.8}$$

where t is the time, P, ET and R are the basin-averaged precipitation, evapotranspiration and runoff, and TWSC is the terrestrial water storage change (Ramillien *et al.*, 2006). It is seen that the performance of a certain model is connected with its differently simulated runoff. At the Amazon basin, the comparably large runoff simulated from MPI-HM also leads to smaller variability in TWS, which is also shown at the Orinoco basin. At the Mekong basin, the larger amplitude in TWS from JSBACH compared with GRACE is related to the apparently small amplitude in its runoff.

4.3.4 Snow-dominated catchments

As highlighted in Section 4.3.2, models perform quite differently at high latitudes of the Northern Hemisphere (cold zone), which are generally dominated by snow. Especially JSBACH and MPI-HM show large differences in the TWS when compared with GRACE. We focus here on four basins in this area, Lena, Yenisei, Ob and Yukon, and look into the mean monthly time series of the TWS and its different components (Fig. 4.11). For LSDM and MPI-HM, subsurface water only includes the water storage in the root zone, while for WGHM and JSBACH, both root zone and deep layer water storage are included. LSDM and WGHM

show the smallest phase differences with GRACE in terms of TWS, while the other two exhibit negative phase shifts. The subsurface water variations from WGHM and LSDM have very similar patterns, with an apparent peak usually in May. The phases of the snow water time series from LSDM and WGHM are also quite close, but LSDM always has a slightly larger amplitude. Since the two use the same snow scheme (degree-day method), this is certainly related to the different model parameters or sub-grid representation schemes. The surface water storage from these two models are sometimes different. For the Ob River, for instance, the different surface water storage also leads to the poor performance of WGHM in terms of TWS when compared with GRACE. The snow variations from LSDM and MPI-HM are almost identical to each other. However, the different subsurface and surface water simulated by MPI-HM causes a bad timing of the TWS peaks. For the Lena basin, although the snow variations from LSDM, WGHM and MPI-HM are quite close, MPI-HM simulates almost no surface water variations, which leads to a poor agreement of TWS with GRACE estimates. For JSBACH, there is already a large phase difference in the snow storage, which is mainly due to the poor capture of the phase of the snow accumulation and onset of melting. This could be caused by the specific snow scheme applied by JSBACH. Yukon, however, is quite different from the other snow-dominated basins. Here, all the models underestimate the annual amplitude of TWS when compared with GRACE. Since the basin-average TWS error from GRACE at Yukon is 1.19 cm and much smaller than the discrepancies between GRACE and the models (Table 4.2), it could be the case that all models fail to represent certain hydrological processes, or that our GRACE TWS errors are too optimistic here since the re-scaling errors are also estimated from a hydrological model ensemble. In addition, Seo *et al.* (2006) found also large TWS errors at Yukon basin and suggested that the atmosphere and ocean tidal and non-tidal de-aliasing errors might be a problem in this area. Investigating those discrepancies in full detail, however, is beyond the scope of our present paper and will be left open for future study.

4.3.5 Dry catchments

We also focus on four catchments in the dry zone, which are characterized by annual precipitation lower than annual potential evapotranspiration (McKnight & Hess, 2000). For the Nile and Niger basins, the subsurface water is the main contributor to the TWS changes (Fig. 4.12). The TWS variations from JSBACH and MPI-HM show a quite similar annual cycle when compared to GRACE. MPI-HM generally exhibits a larger amplitude in simulated subsurface water and TWS. WGHM deviates considerably with a much smaller amplitude and a large phase shift in the subsurface water. The simulated surface water from WGHM

brings TWS slightly closer to that from GRACE. LSDM, however, performs differently in these two basins. In the Nile basin, although the subsurface water from LSDM is consistent with JSBACH and MPI-HM, the simulated surface water variations lead to a higher amplitude of TWS variations when compared with GRACE. In Niger, LSDM performs quite closely to WGHM, but with a slightly larger amplitude. All models tend to perform poorly in terms of TWS when compared with GRACE in the Indus basin. We note a comparably low SNR (2.2 cm) for the GRACE estimated TWS here, which is mainly contributed by the large leakage error at this basin (Zhang *et al.*, 2016). In addition, the Indus basin is not only subject to large-scale groundwater depletion from intensive irrigation, but is also affected by snow melting and glacier melting from Himalaya. Here, the subsurface water simulated by the models already shows large discrepancies. As in other basins affected by snow dynamics, JSBACH also fails to capture the snow variations properly. MPI-HM performs poorly in simulating the surface water, with a delayed dynamics which leads to a preceded annual cycle. At the Huang He basin, the subsurface water from LSDM, WGHM and JSBACH as the main contributors to the TWS show similar annual variations to GRACE, while MPI-HM has a much larger amplitude. The surface water, however, is simulated differently by LSDM and WGHM, which consequently leads to different TWS variations.

4.4 Summary

We validate TWS variations simulated by four different global hydrological models with monthly GRACE gravity data. All the models are forced with the same WFDEI meteorological data set to exclude the effect of meteorological forcing on the models. Four statistical metrics focusing on different aspects of model performance compared with GRACE have been applied. In addition, time series of TWS variations from GRACE and models are investigated, where different water storage components from models are shown as well.

At certain basins like the Danube, Tocantins, Columbia, Ganges, Mekong, and Amazon, all numerical models show good agreement with GRACE. However, models still perform quite differently at many other basins, even though forced with the same meteorological data set. At the Nile, Indus, Murray and Great Artesian basins, large TWS errors and low SNR are found, which suggests a major contribution from GRACE errors to the differences. A good capture of annual amplitude and phase at most basins leads to high values of explained variance in many basins for LSDM. However, serious problems are also found in the same model run in some central Africa basins, like the Nile and Zaire, where TWS simulated by LSDM exhibits unusual large inter-annual variations. WGHM performs generally well

in tropical and cold regions, but rather poorly in the temperate zone. JSBACH and MPI-HM show large discrepancies with GRACE at the basins at high latitudes of the Northern Hemisphere.

Model performance is also investigated in some snow-dominated and dry catchments in more detail through time series comparison. The poor performance of JSBACH and MPI-HM in snow-dominated regions is mainly related to negative phase shifts compared to GRACE. MPI-HM simulates identical snow variations to LSDM; however, the different simulations of subsurface water and especially surface water still lead to different TWS variations in snow-dominated regions. Despite the missing surface water component, the simulated snow variations in JSBACH already show smaller amplitude and negative phase differences compared with all the other models. This could be related to the fact that JSBACH simulates snow in a more physical way based on energy balance, which is totally different from the degree-day method applied by all the other models. The comparably better agreement of LSDM and WGHM with GRACE in terms of TWS in these snow-dominated basins is partly caused by the realistic surface water component represented by these two models. In the dry catchments, the impact from AET on TWS is relatively strong. The smaller AET from MPI-HM also leads to better agreement with GRACE, whereas LSDM shows large differences with GRACE in terms of TWS, especially at some dry basins in central Africa, partly due to the overly simple evaporation scheme. PET is simulated using a superior parametrization by MPI-HM, while LSDM still applies the traditional Thornthwaite method based solely on air temperature. The groundwater considered by WGHM also has some impact on the simulated TWS, especially at basins such as Tocantins, Mekong, Niger and Mississippi. At the Yukon basin, we found the bad performance of all models in terms of TWS when compared with GRACE, which could be due to the effects of atmospheric and oceanic de-aliasing errors not further discussed in our current study. In future, we would like to assess all possible errors of GRACE TWS through investigation of simulated GRACE-type gravity field time series (Flechtner *et al.*, 2016) based on realistic orbits and instrument error assumptions as well as background error assumptions out of the updated ESA Earth system model (Dobslaw *et al.*, 2015, 2016), which we believe will further help to explain the discrepancy between global models of the terrestrial water cycle and GRACE satellite observations.

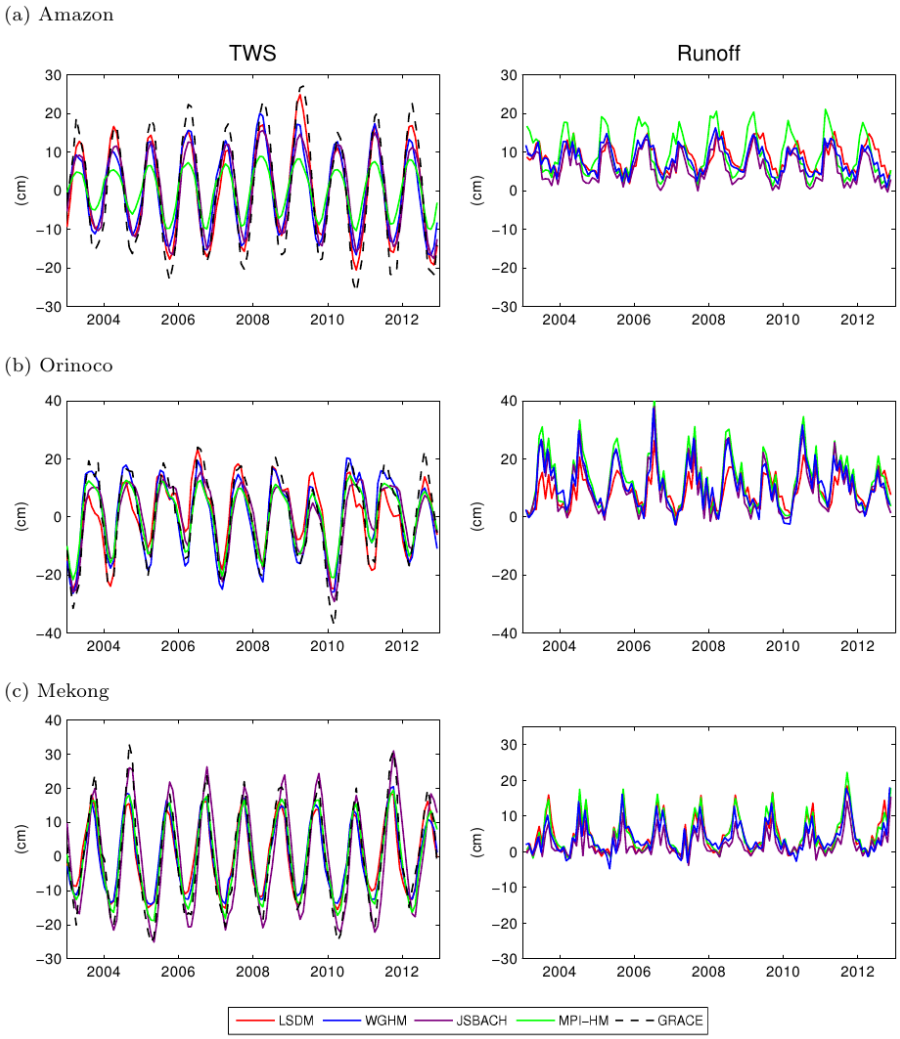

Figure 4.10: Time series of TWS (left) from GRACE and models and model simulated runoff time series (right); each for three different catchments in tropical zone: Amazon, Orinoco, and Mekong.

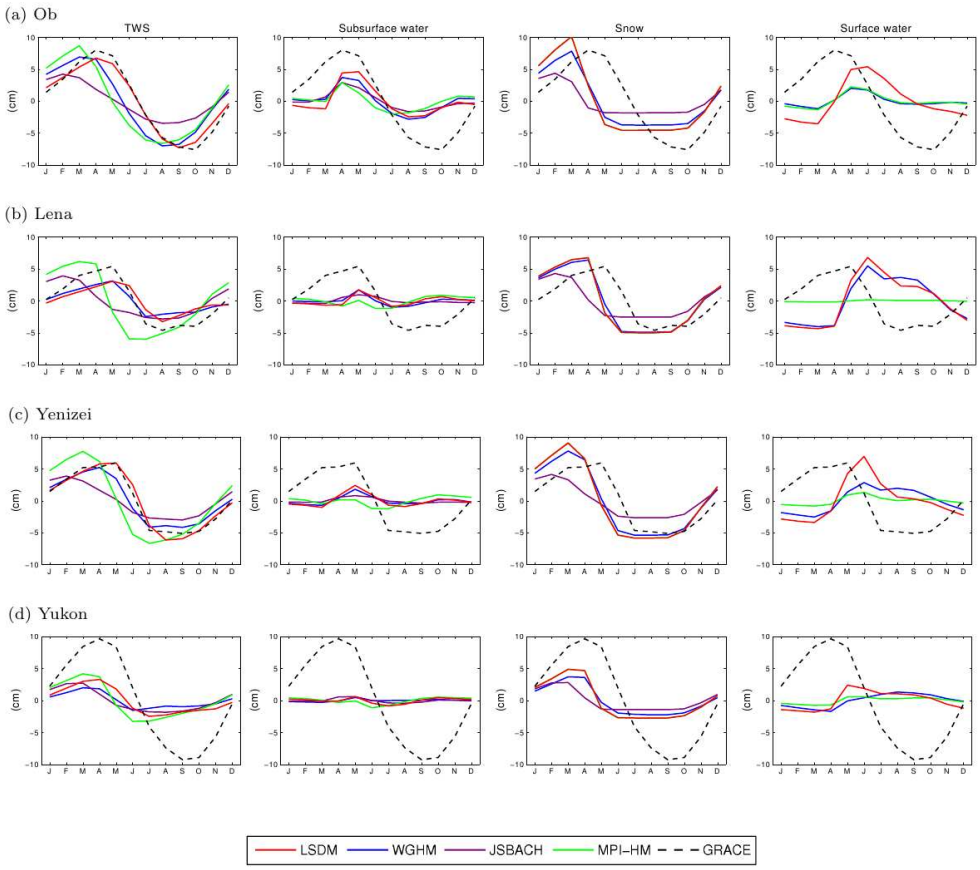

Figure 4.11: Mean monthly time series of TWS (first column) and the individual storage contributions from subsurface water (second column), snow water equivalent (third column) and surface water (fourth column); each for four snowy catchments: Ob, Lena, Yenizei and Yukon. TWS from GRACE (dashed line) has been included into every sub-figure for reference.

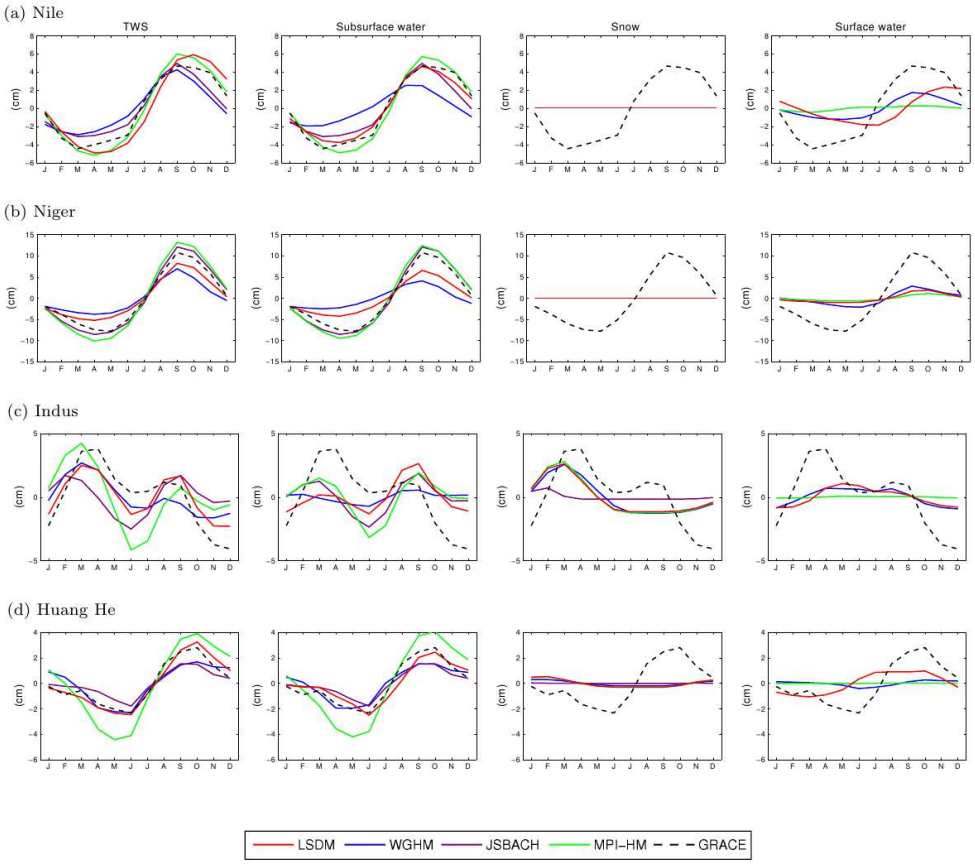

Figure 4.12: Mean monthly time series of TWS (first column) and the individual storage contributions from subsurface water (second column), snow water equivalent (third column) and surface water (fourth column); each for four dry catchments: Nile, Niger, Indus and Huang He. TWS from GRACE (black line) has been included into every sub-figure for reference.

Chapter 5

Validation of MPI-ESM Decadal Hindcasts

5.1 Decadal predictions and MPI-ESM hindcast experiments

5.1.1 Decadal predictions

Near term climate predictions, ranging up to one decade in the future, are also called decadal predictions. As a new endeavor in climate change, it can provide information about the climate state over the next few years, which is very important for policy and decision makers. For instance, it offers potential to assess the probability of extreme events, like drought, flood, wildfire, so that comprehensive planning can be made ahead.

The time horizon of decadal prediction lies between seasonal-to-interannual predictions and long-term projections (Meehl *et al.*, 2009), and it thus combines some features of the two time scales. Numerical weather forecasts predict the evolution of the atmospheric circulation for the following days, which requires the information of the current or initial state of the system, predominantly the atmosphere and land surface. On longer time scales climate predictions rely on history and future evolution of natural and anthropogenic forcings (Smith *et al.*, 2016). Similar to seasonal-to-interannual predictions, decadal predictions also need information about the current state. It, however, relies on the evolution of the oceanic circulation over the next few years, as the oceans are evolving more slowly and have longer "memory" than the atmosphere. Thus, the decadal predictions depend not only on how it captures the current state of the system which is addressed as "initial value problem", but also how well it captures the anthropogenically-forced and naturally-forced climate changes, referred as "Boundary condition problems" (Fig. 5.1; Meehl *et al.*, 2009).

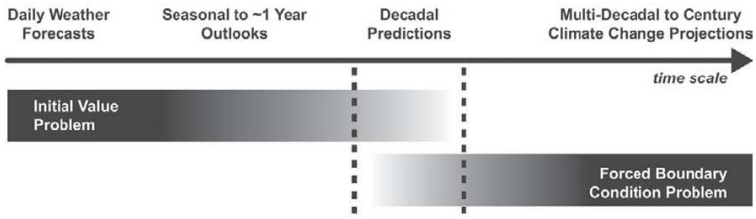

Figure 5.1: Schematic illustrating progression from initial value problems with daily weather forecasts at one end, and multi-decadal to century projections as a forced boundary condition problem at the other, with seasonal and decadal prediction in between. (Meehl *et al.*, 2009)

Generally, decadal predictions are produced using the same models as for long-term climate projections. A climate model is a simplified mathematical representation of the climate system, and it includes the four fundamental elements of it (atmosphere, ocean, land and cryosphere). The climate models are normally initialized with the observed ocean temperature and salinity fields, and some with sea-ice and atmospheric variables at the starting time of the decadal simulation. With this initialization, the simulations start from a climate state that better matches the current phase of the observed climate. To assess the predictive skill of decadal experiments, retrospective forecasts, named "hindcasts", are constructed from the historical observations.

5.1.2 MPI-ESM hindcast experiments

The current ongoing German research project MiKlip (Mittelfristige Klimaprognosen) is dedicated to develop a climate forecast system that is able to provide skillful decadal predictions for up to a decade ahead. This system is based on the Max Planck Institute Earth System Model (MPI-ESM). As an Earth system model (ESM), it includes processes of the atmosphere, the land and the ocean simulated by different model components. An overview of these modules is given in Giorgetta *et al.* (Fig. 5.2; 2013). The dynamical atmospheric part of the MPI-ESM is simulated by the general circulation model European Center-Hamburg Atmosphere Model version 6 (ECHAM6, Stevens *et al.*, 2013), while the Max Planck Institute Ocean Model (MPIOM, Jungclaus *et al.*, 2013) simulates the ocean. Both modules are coupled through an interface to the Jena Scheme for Biosphere Atmosphere Coupling in Hamburg (JSBACH, Reick *et al.*, 2013; Hagemann *et al.*, 2013) and the Hamburg Ocean Carbon Cycle Model (HAMOCC, Ilyina *et al.*, 2013), respectively, which is performed by the Ocean Atmosphere Sea Ice Soil (OASIS, Valcke, 2013).

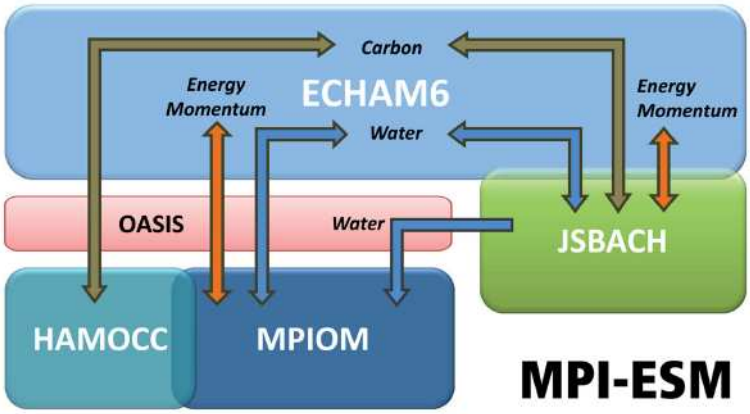

Figure 5.2: Schematic view of MPI-ESM: Colored boxes show the model components: ECHAM6 is the atmospheric general circulation model, which is directly coupled to the JSBACH land model that describes physical and biogeochemical aspects of soil and vegetation. MPIOM is the ocean general circulation model, which includes the HAMOCC model for the marine biogeochemistry. OASIS is the coupler program, which aggregates, interpolates, and exchanges fluxes and state variables once a day between ECHAM6+JSBACH and MPIOM+HAMOCC. The coupler exchanges fluxes for water, energy, momentum, and CO2 (Giorgetta *et al.*, 2013).

Numerous long-term experiments have been performed with the MPI-ESM following the 5th Phase of the Coupled Model Intercomparison Project (CMIP5) experiment protocol (Taylor *et al.*, 2012). The long-term uninitialized historical simulation starts in 1850 and goes until 2005 driven with prescribed natural and anthropogenic forcings, like solar and volcanic variability and anthropogenic greenhouse gas concentrations and aerosols. For the subsequent years, projections following the RCP4.5 emission scenario are applied (Giorgetta *et al.*, 2013). The historical simulations include ten ensemble members that are randomly initialized in 01/1850, realized by using lagged initialization for different start dates. In addition to the long-term experiments, decadal climate predictions (CMIP5) are also available.

Within MiKlip, there are three decadal prediction hindcast experiments, baseline-0 (b0), baseline-1 (b1) and prototype (pr). They mainly differ from each other in their employed initialization technique and the ensemble size (Müller *et al.*, 2014). The b0 hindcast is analogue to the initialized CMIP5 hindcast. The anomaly technique (Pierce *et al.*, 2004; Smith *et al.*, 2013) is applied to initialize the b0 hindcasts with the oceanic temperature and salinity anomaly fields from MPIOM forced with NCEP/NCAR atmospheric reanalysis data (Kalnay *et al.*, 1996). The simulations were initialized for the period 1961 to 2012, each simulating a decade with three ensemble members. An additional set of 7 ensemble members is initialized every five years. Two different model configurations LR/MR are implemented,

that differ from each other not only in vertical resolution of atmospheric part but also in horizontal resolution of the ocean (Giorgetta *et al.*, 2013). The atmospheric component of b0-LR is ECHAM6 in T63L47-resolution (approximately 210 km horizontal grid spacing at the equator), while the ocean is represented by MPIOM with GR15L40-resolution (grid spacing ranging from approximately 15 km around Greenland to 185 km in the tropical Pacific) (Kruschke *et al.*, 2015). The 47 vertical pressure levels of the atmosphere in the b0-LR is doubled to 95 in MR. For the ocean, b0-MR uses a tri-polar grid (TP) with a horizontal resolution of approximately 0.4 ° at the equator (Giorgetta *et al.*, 2013). The ocean in the b1 simulations is also initialized using the anomaly technique, but with ocean reanalysis data from the ocean reanalysis system 4 (ORA-S4) from ECMWF (Balmaseda *et al.*, 2013). In addition, the atmospheric components in b1-LR is initialized with full-field parameters as surface pressure, temperature, vorticity and divergence from ERA-Interim (Dee *et al.*, 2011). An ensemble of 10 members is lagged initialized each year. The b1-MR experiment is similar with the LR version except for the higher resolution in ocean and atmosphere as mentioned above and its number of ensemble members (five members). The prototype prediction system initializes the atmosphere the same way as b1, but applies full field method to initialize the ocean with temperature and salinity fields taken from ORA-S4 and German contribution to Estimating the Circulation and Climate of the Ocean (GECCO2) (Köhl, 2015). The ensemble size is increased to 30, with 15 ensemble members based on ORA-S4-initialized simulations and additional 15 members initialized with GECCO2. Both b1 and pr are initialized yearly for the period from 1961 to 2013.

5.1.3 Verification

Decadal predictions take into account both natural variability and forced climate change. They requires information of the atmosphere, oceans, cryosphere, and land surface so that the coupled general circulation models can be best initialized. Meanwhile, these initial states are not only affected by the natural variability, but by the anthropogenic forcing as well. There are, however, no comprehensive observations of the ocean and the climate models still have large uncertainties. These all make the decadal prediction quite challenging.

Due to these difficulties, the decadal prediction research is still experimental. The decadal prediction experiments can be used to compare the prediction quality across the different experiments and assess what predictions should be attempted and how it should be done. Goddard *et al.* (2013) introduced a framework to verify decadal hindcast experiments, which mainly focuses on these research questions: 1) Does initialization of the hindcasts with observations of the slowly-varying components of the climate system result in more accurate

near-term predictions in comparison with uninitialized projections? and 2) Is it plausible to use the ensemble spread of the prediction model to represent forecast uncertainty on average?

5.1.3.1 Preprocessing

Before employing the set of metrics to assess prediction quality, both model data and observations have to be re-arranged to be consistent. First, the ensemble averages are calculated as the mean of all ensemble members H_{ijk} that differ only in their initial conditions.

$$H_{ij} = \frac{1}{\mathrm{N1}} \sum_{k=1}^{\mathrm{N1}} H_{ijk} \tag{5.1}$$

where k=1, N1 is the set of ensemble members, run at each initial time i=1, N2, and extend over a prediction range of j=1, N3. It is known that the model tends to drift away from its initial state which is close to the observations toward its own model climate due to the problems with the model equations, numeric schemes and parameterizations, leading to bias in the forecasts (Zhang, 2011). The drift can be calculated as the climatological averages of a series of initialized hindcasts over all starting dates (Smith *et al.*, 2013), or as climatological average for each time period from the hindcasts from the corresponding period (García-Serrano & Doblas-Reyes, 2012), or as a function of lead-time and calendar month based on a set of hindcasts. The lead-time dependent prediction anomaly is estimated by subtracting from every lead month t_j (with j = 1 ;...; N2) of a given hindcast experiment H_i (i = 1 ;...; N2) the climatological mean of the same lead months from all other experiments:

$$\Delta H_{ij} = H_{ij} - \frac{1}{\mathrm{N2} - 1} \sum_{p \neq i} H_{pj} \tag{5.2}$$

This procedure is then applied to all other lead months of all experiments. The long-term simulations and observations are handled differently, where the climatological mean covering the whole time range is subtracted.

As small-scale variability in spatial and temporal scales could contaminate the larger-scale climate variability, noise could be introduced, which reduces the prediction skill. On temporal scale, different scales: year 1, years 2-5, years 6-9 and years 2-9, are recommended to assess the skill for different lead times and temporal averaging. The temporal average between the initial year, Y_1 and the final year Y_{N3} of a certain hindcast is calculated as:

$$H_i = \frac{1}{(Y_{\mathrm{N3}} - Y_1 + 1)} \sum_{j=Y_1}^{Y_{\mathrm{N3}}} H_{ij} \tag{5.3}$$

The long-term simulations and observations are sampled as the average of the same respective years as is done for the hindcasts. Spatial smoothing is sometimes applied before verification to reduce the unpredictable grid-scale noise. On the spatial scale, $5° \times 5°$ is suggested to be used for smoothing precipitation, and $10° \times 10°$ for temperature, since precipitation is derived from more localized processes (Räisänen & Ylhäisi, 2011).

5.1.3.2 Evaluation metrics

To solve the first question of whether the initialization results in more accurate predictions, a deterministic metric, the mean squared skill score (MSSS), is applied. The MSSS is calculated from a set of hindcasts H_i and the observations O_i via the mean squared error (MSE):

$$MSE = \frac{1}{\mathrm{N2}} \sum_{i}^{\mathrm{N2}} (H_i - O_i)^2 \tag{5.4}$$

where H_i and O_i are both anomalies for which the mean bias has already been removed. The MSSS of a prediction with respect to the reference, like the climatological average $\bar{O}$ can then be written as:

$$MSSS(H, \bar{O}, O) = 1 - \frac{MSE_H}{MSE_{\bar{O}}}. \tag{5.5}$$

The MSSS can be also defined using the uninitialized predictions (P) instead of the observations as reference:

$$MSSS(H, \bar{P}, O) = 1 - \frac{MSE_H}{MSE_{\bar{P}}}. \tag{5.6}$$

A perfect MSSS has the value 1, where the mean squared error of the targeted predictions is 0. Accordingly, a positive value indicates the predictions H is more skillful than the reference predictions, whereas a negative MSSS means just opposite.

Since the decadal prediction is a probabilistic forecast, it is important to assess the range of possibilities for individual predictions. The probabilistic skill is evaluated through the continuous ranked probability skill score (CRPSS), which can also be used to solve the second question of whether the ensemble members spread represents the possibilities of the predictions. The CRPSS is derived from the continuous ranked probability score (CRPS), which quantifies the difference between the predictive cumulative distribution functions (cdf) of the hindcasts and the cdf of the observation (Goddard *et al.*, 2003):

$$CRPS(H_{ik}, O_j) = \int_{-\infty}^{+\infty} (\mathcal{G}(H_j) - \mathcal{H}(O_j))^2 dy \tag{5.7}$$

where $\mathcal{H}$ represent the Heaviside function assuming perfect observations and $\mathcal{G}$ represents the Gaussian distribution centered on the mean H_j and ensemble variance about that mean, σ_H^2. Taken the $CRPS_F$ as the hindcast distribution, and the $CRPS_R$ as the reference distribution

$$CRPSS = 1 - \frac{\sum_{i=1}^{\mathrm{N2}} CRPS_{F_i}}{\sum_{i=1}^{\mathrm{N2}} CRPS_{R_i}}, \tag{5.8}$$

shows the way to estimate the CRPSS. To solve the question mentioned above, the hindcast variance is calculated as:

$$\sigma_F^2 = \frac{1}{\mathrm{N2}} \sum_{i=1}^{\mathrm{N2}} \frac{1}{\mathrm{N1}-1} \sum_{k=1}^{\mathrm{N1}} (H_{ik} - H_i)^2, \tag{5.9}$$

while the variance for the reference distribution is expressed as:

$$\sigma_R^2 = \frac{\sum_{i=1}^{\mathrm{N2}} (H_i - O_i)^2}{\mathrm{N2}-2}. \tag{5.10}$$

For the skill verification, it is also important to evaluate the statistical significance of scores. Here, a non-parametric block bootstrap approach (Wilks, 2011; Goddard *et al.*, 2013) has been applied. Firstly, a re-sampled set of $p = 1, M$ initialized and uninitialized simulations is made. Accordingly, the $\tilde{H}_{IK}$ is sampled by first randomly selecting the new start times from the set of years and then setting the new N1 ensemble members from each start time. This new sampled bootstrapped hindcasts can then be applied to derive different score distributions. Take MSSS score for instance, the distribution of MSSS values can be obtained from the bootstrap-generated time series from $\widetilde{MSSS}(k)$. The p value for the test is represented by the fraction of values that are negative. If p is smaller than or equal to the selected significant level α, the score of this grid cell is considered significant for the $(1-\alpha) \times 100\%$ confidence level.

5.2 Deterministic skill investigation of MPI-ESM decadal hindcasts with TWS from GRACE

Published as: Zhang, L., Dobslaw, H., Dahle, C., Sasgen, I., and Thomas, M. (2015). Validation of MPI-ESM decadal hindcast experiments with terrestrial water storage variations as observed by the GRACE satellite mission, Meteorol. Z., doi: 10.1127/metz/2015/0596.

5.2.1 Introduction

Reliable predictions of near-future changes in the Earth's climate beyond the seasonal timescale would be highly valuable for the human society. Since adaption to and mitigation of climate change require far-reaching decisions many years before any measure might be effective, reliable estimates on low-frequency climate variability including near-term trends with lead time of a few years would be very beneficial for policy management and decision-making on climate change investments. This includes, for example, predicting the probability of extreme events as demonstrated in skillful multi-year predictions of Atlantic hurricane frequency (Smith *et al.*, 2010), but also near-surface air temperatures and the global water cycle.

The assessment of forecast skill of any decadal prediction system is typically based on validating extensive sets of hindcast experiments, where observations are already available (Smith *et al.*, 2007; Keenlyside *et al.*, 2008). Besides atmospheric and oceanic state quantities, also land surface conditions are increasingly recognized as influential for the evolution of climate in particular on the longer time scales. Water stored on the continents does not only affect the atmospheric circulation by means of surface albedo changes and thermal isolation due to snow cover, but also influences evaporation (Koster *et al.*, 2004a; Meehl *et al.*, 2009; Seneviratne & Stöckli, 2007). Further, groundwater has a potentially large impact on the low-frequency climate variability by means of its contributions to soil moisture re-charge (Bierkens & van den Hurk, 2007).

For a thorough skill assessment of any decadal prediction system it is therefore important not to focus solely on traditionally well-observed quantities as air temperatures or precipitation, but to take additionally many more types of observations into account (Mahmood *et al.*, 2010). The satellite mission Gravity Recovery and Climate Experiment (GRACE; Tapley *et al.*, 2004b) provides since 2002 monthly snap-shots of the time-variable global gravity field of the Earth. Over the continents, gravity variations are primarily related to changes of water stored as snow, soil moisture, surface water, groundwater, and biomass. Since mass anomalies affect gravity measured by GRACE independently of a surface exposure, the GRACE experiment is the only realization of a remote sensing technique that is able to provide estimates of water masses integrated vertically from the surface down to the deep aquifers. The sum of the water masses in all those different hydrological storage components observed by GRACE is typically denoted as terrestrial water storage (TWS). TWS can be viewed as a measure of the vertical and lateral water fluxes in time and is therefore characterized by distinct low-frequency variability. This low-frequency character makes TWS ideally suit for decadal climate prediction model validation efforts. However, even though TWS estimates

from GRACE are available for more than one decade up to now, the currently available knowledge on the overall accuracy of the GRACE-based TWS in particular on periods longer than seasonal is still limited, since it relies so far primarily on the assessment of hydrological model data (Syed *et al.*, 2008b) or the analysis of combined atmospheric-terrestrial water balances (Seneviratne *et al.*, 2004).

In this thesis, we examine three different sets of decadal hindcasts performed with the coupled Earth System Model from Max-Planck-Institute for Meteorology (MPI-ESM) within the German research project "Mittelfristige Klimaprognosen" (MiKlip) (Müller *et al.*, 2012; Pohlmann *et al.*, 2013). We follow the verification framework proposed by Goddard *et al.* (2013) to assess the prediction quality of the MPI-ESM decadal hindcasts against terrestrial water storage anomalies from GRACE. Other satellite observations like the upper troposphere and lower stratosphere region (UTLS) temperature and satellite-retrieved cloud parameters have also been applied to evaluate the same MPI-ESM decadal hindcast experiments (Schmidt *et al.*, 2015; Spangehl *et al.*, 2015). The structure of our paper is arranged as follows: the processing of the GRACE satellite observations is outlined in Sec. 5.2.2. Subsequently, Sec. 5.2.3 gives a brief summary on the available hindcast experiments, whereas Sec. 5.2.4 sketches the validation approach applied. Results for deterministic skill scores are presented in Sec. 5.2.5, followed by some conclusions in the final section of this paper.

5.2.2 GRACE satellite observations

We use monthly GRACE RL05 Level-2 products from January 2003 to December 2011 processed at GFZ (Dahle *et al.*, 2012) and estimate globally gridded and re-scaled TWS estimates from GRACE following the same procedure as introduced in Section 3.3. Then the gridded TWS variations are aggregated into 5° averages as suggested by Goddard *et al.* (2013), and a mask is applied that removes ocean, small islands, and areas where the standard deviation of TWS as seen by GRACE is smaller than 1 cm of equivalent water height (eq.w.h.).

The time-variability in TWS as seen by GRACE is primarily dominated by the seasonal cycle (Fig. 5.3a). We note standard deviation values of up to 15 cm eq.w.h. in tropical areas as, e.g., in the Amazon catchment, the Congo basin, and the Indian Monsoon area. Variabilitiy is substantially smaller at higher latitudes, but still reaches standard deviation values of up to 10 cm eq.w.h. in snow-dominated regions of North America and Eurasia. Calculating standard deviations from yearly averages (Fig. 5.3c) in order to focus on differences between individual years only indicates much smaller variability of up to 5 cm eq.w.h. It is interesting to note that for the annual means the TWS variability at moderate latitudes is equally high as in tropical catchments.

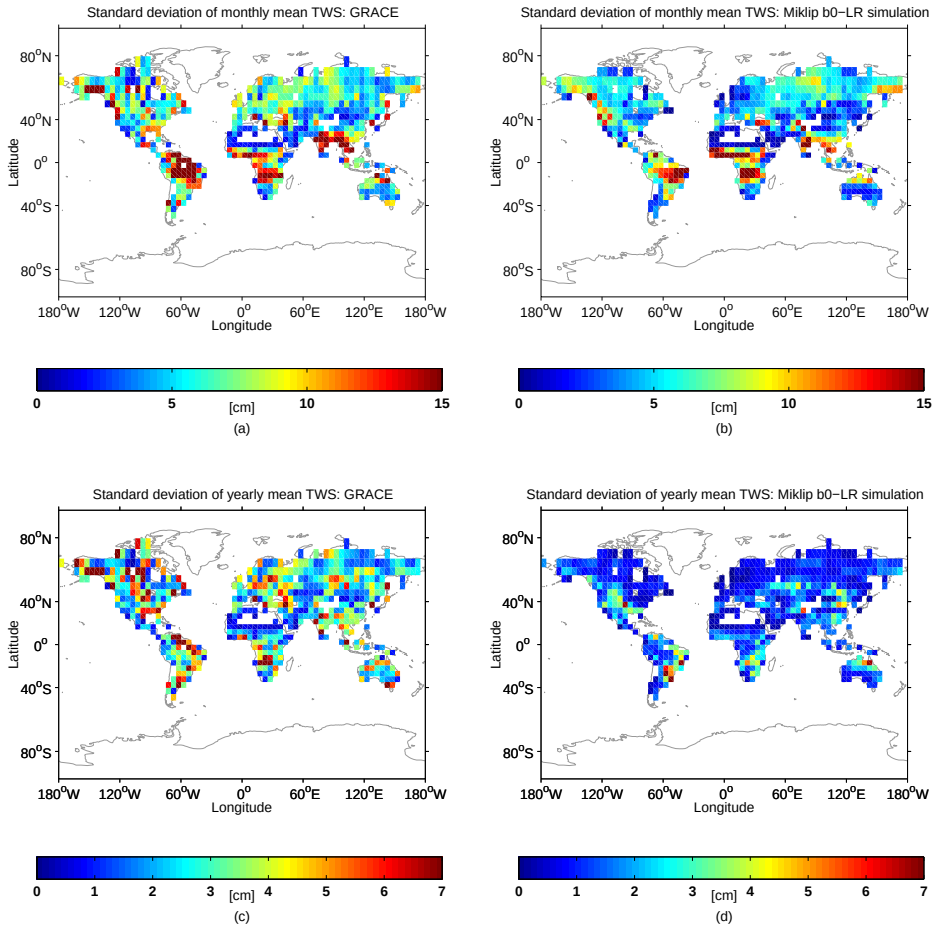

Figure 5.3: Standard deviation of monthly (top) and yearly averaged (bottom) TWS anomalies from GRACE (left) and the ensemble mean of a decadal hindcast experiment b0-LR (right) for lead year 1. Both data-sets are available from January 2003 to December 2011 and are aggregated here into 5° grid cells.

5.2.3 Decadal hindcasts from MPI-ESM

I apply three different ensemble sets of decadal hindcasts that are available from two different versions of the MPI-ESM coupled climate model. The low resolution (LR) model variant has been used already within the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor *et al.*, 2012) and includes the oceanic component MPIOM (Jungclaus *et al.*, 2013) discretized on a 1.5° curvilinear grid with 40 layers in the vertical, the atmospheric model ECHAM6 (Stevens *et al.*, 2013) at T63 (1.875°) horizontal resolution with 47 vertical levels that reach up to 0.1 hPa in the upper stratosphere, and the land surface model JSBACH (Reick *et al.*, 2013; Hagemann *et al.*, 2013). A second variant of MPI-ESM with medium resolution (MR) has a finer horizontal resolution in the ocean (0.4° curvilinear grid), and 95 levels in the atmosphere.

For both model variants, historical runs in line with the CMIP5 protocol are available until 2006. For the subsequent years, projections following the RCP4.5 emission scenario are used. In the remainder of this thesis, we denote those runs as "uninitialized projections". In order to cover the same time span, GRACE observations in the period 2003 - 2011 have been applied, and the following sets of decadal hindcasts are considered in the subsequent analysis:

(i) baseline0 at low resolution (b0-LR): three ensemble members initialized every year between 1994 and 2000, ten ensemble members initialized every year between 2001 and 2010;

(ii) baseline1 at low resolution (b1-LR): ten ensemble members initialized every year between 1994 and 2010; and

(iii) baseline1 at medium resolution (b1-MR): five ensemble members initialized every year between 1994 and 2010.

For all model runs, the standard output variables "total soil moisture content" (mrso) and "surface snow amount" (snw) are taken as the dominant contributors to the terrestrially stored water; the sum of both is contrasted against GRACE-based TWS in the remainder of this thesis. Note that both surface water and deep aquifers are likely to contribute to TWS as well, their effects have been, however, omitted in this assessment, since none of them is properly represented by MPI-ESM. We are primarily interested in assessing the hindcast quality in relation to the lead time, thus we re-sort the data and build lead year time-series for the ensemble means of each ensemble set of hindcasts.

Exemplarily, variability of TWS as simulated within the b0-LR hindcast experiment is depicted in Fig. 5.3b. In general, signal magnitudes and spatial patterns fit well to the GRACE observations. Boundaries between different hydro-climates are more sharply defined in the

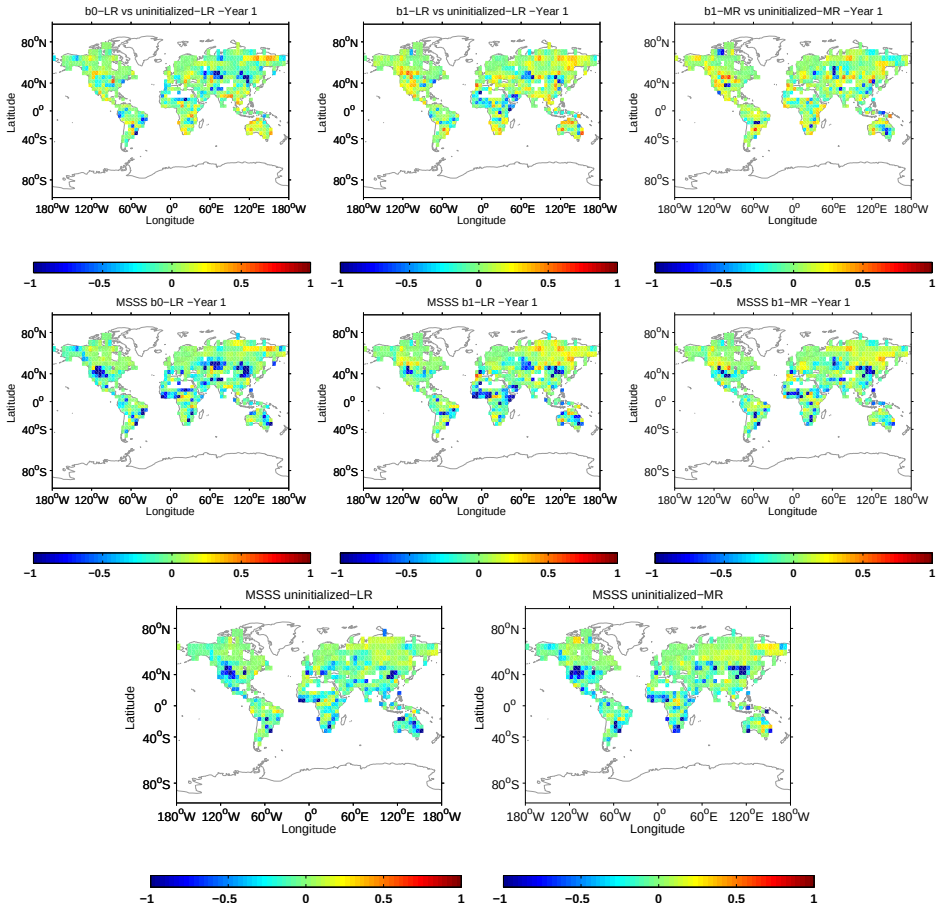

Figure 5.4: Mean squared skill score (MSSS) for three different decadal hindcast experiments within the MPI-ESM with respect to GRACE-based TWS observations for the period 2003 - 2011. Top row: MSSS of initialized hindcasts vs. uninitialized projections for lead year 1 of the experiments b0-LR (left), b1-LR (middle), and b1-MR (right). Middle row: MSSS of initialized hindcasts vs zero anomaly forecasts for lead year 1 of the experiments b0-LR (left), b1-LR (middle), and b1-MR (right). Bottom row: MSSS of uninitialized projections vs. zero anomaly forecasts of the uninitialized-LR (left), and the uninitialized-MR run.

model results, suggesting that spatial leakage is still present in the post-processed and re-scaled GRACE results. For the standard deviation of annual means, however, we note substantially lower variability compared to the GRACE data in almost all regions of the world, and the magnitudes typically reach only half the amount of the observations (Fig. 5.3d).

5.2.4 Validation approach

By using TWS from GRACE as the reference observations, we are going to assess how skillful the three different sets of hindcasts from the MPI-ESM are in terms of predicting anomalous terrestrial water storage. By posing this validation question, we attempt to investigate whether the models provide information about water availability below the long-term mean (i.e., which might lead to drought conditions) or above it (i.e., which might lead to flood conditions). Since the trends in TWS observed by GRACE cannot be exclusively assigned to natural climate variability, but are also due to, for instance, solid earth geophysics and anthropogenic groundwater extraction from deep aquifers, the average trends of TWS over the observation period are removed from both the observations and hindcasts. To exclude also seasonal effects, which are not in the focus of a decadal climate prediction system, we confine ourselves to the assessment on annual averages only. Finally, both model and observational data-sets are aggregated onto a regular 5° grid as used also in Goddard *et al.* (2013). As a deterministic metric, we use the mean-squared skill score (MSSS, Eq. 5.5), that compares the mean-squared errors (MSE, Eq. 5.4). This score is also implemented in the central evaluation system of MiKlip as described by Kadow *et al.* (2015). Here, both P_j and O_j are anomalies relative to their respective climatologies calculated over the same data span. Besides the uninitialized projections, we also use "zero-anomaly forecasts" as a reference R in Eq. 5.6. As a deterministic score, the MSE is typically calculated from ensemble mean averages that we obtain from the re-sorted lead year dependent time-series from the different sets of ensemble hindcasts as described in the previous section. Besides showing MSSS for the single lead year 1, average MSSS over lead years 2 - 5, and 6 - 9 are also calculated. Following the non-parametric bootstrapping method described in Appendix 2 of Goddard *et al.* (2013), we also assess the significance of the MSSS of initialized hindcasts b1-MR with respect to uninitialized projections for lead year 1 with three different confidence levels of 90%, 95% and 99% .

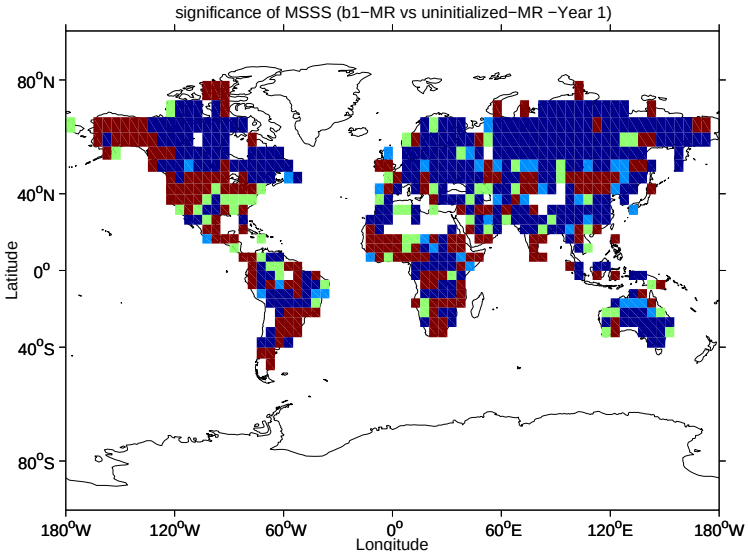

Figure 5.5: Significance of the MSSS of initialized hindcasts b1-MR with respect to uninitialized projections for lead year 1. Red areas shows statistical significance that the MSSS is positive at the 99% confidence level, green indicates the added areas where the confidence level is chosen as 95%, light blue is for 90% and dark blue represents the other areas.

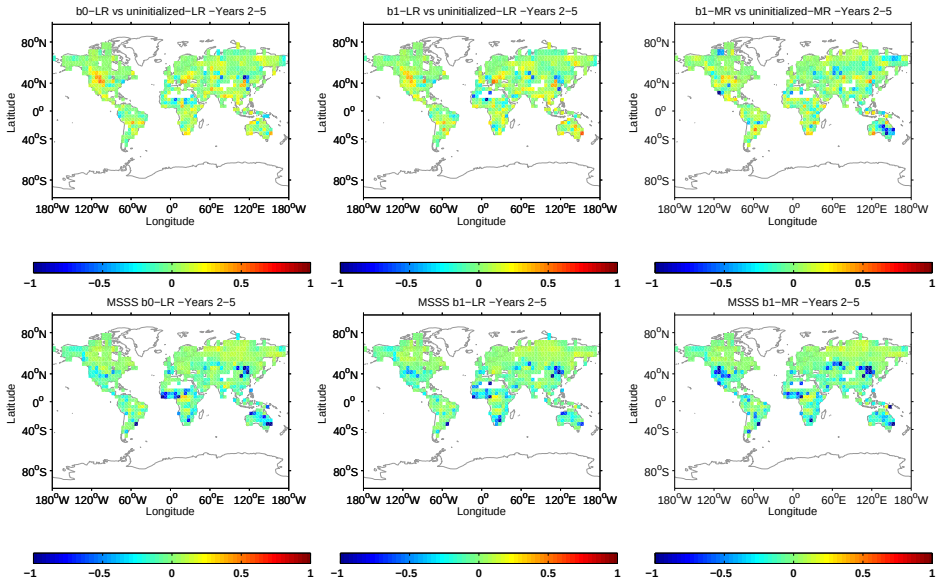

Figure 5.6: Same as Fig. 2, but for lead years 2-5.

5.2.5 Results

With globally gridded TWS anomalies at 5° spatial resolution from GRACE as the reference, MSSS maps are calculated for the three different MPI-ESM decadal prediction ensembles. Focusing on lead year 1 (Fig. 5.4), we note that the scores of the uninitialized projections compared to a zero-anomaly forecast are generally small for the LR and MR model version considered: estimates hardly exceed 0.5 even in isolated regions. Nevertheless, we note a slight improvement of the scores for MR in moderate to high latitudes of the Northern Hemisphere, and also a reduction of the highly negative skills shown in LR in the Southeast-Asian Monsoon regions. When looking into scores of the initialized runs with respect to a zero-anomaly forecast, we find moderate skills in large parts of Siberia in b0-LR that are substantially improved in b1-LR, indicating that the combined atmosphere/ocean initialization strategy implemented in baseline1 is also beneficial to the representation of land-surface processes in MPI-ESM. This improvement is also reflected in scores calculated for the initialized versus the uninitialized projections, that also increase from b0-LR to b1-LR, culminating in the latter in positive skill scores in almost all regions apart from the large deserts, where TWS variations are small and observations are certainly dominated by errors. Skills are also largely positive for b1-MR when comparing predictions to the uninitialized projections. However, since the uninitialized projections already show better skill than the LR projection with respect to a zero-anomaly forecast, there is no obvious improvement compared with b1-LR. We note from the significance figure (Fig. 5.5) that the regions where we discussed changes in the skill scores (e.g., parts of Northern Eurasia, Monsoon areas) indeed show significant skills.

For longer lead time of the predictions, however, skill scores of all three model runs diminish since the influence of the initialization becomes increasingly smaller. Skill scores averaged over lead years 2-5 are generally smaller when calculated against the zero-anomaly forecast (Fig. 5.6), and the distinctness of the two different initialization strategies or model versions are not substantial anymore. With respect to the uninitialized projections, we still note generally positive skills for all three hindcast ensembles considered, but notable differences between baseline0 and baseline1, or LR and MR are no longer evident, and scores are regionally much less coherent than for lead year 1. Generally the same holds true for the averaged scores over lead years 6- 9 (Fig. 5.7). Although still positive scores are obtained for the initialized forecasts when compared to uninitialized projections, those are typically only slightly larger than zero and do show very little regional coherence.

To analyze the time-variability of the simulated terrestrial water storage for selected regions in more detail, we average TWS from both GRACE and the different model runs for two

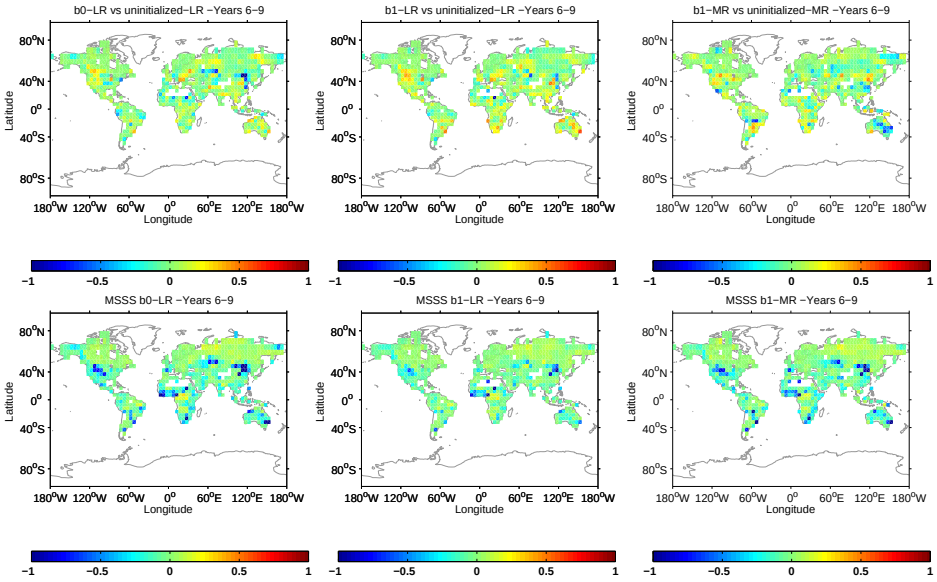

Figure 5.7: Same as Fig. 2, but for lead years 6-9.

differently located discharge basins: The Lena catchment in Central Siberia whose climate is generally classified as "snow dominated, fully humid with cool summer (Dfc)" according to the Köppen-Geiger Climate Classification (Kottek & Grieser), and the Nelson River catchment in North America with a slightly different hydro-climate of "snow dominated, fully humid with warm summer (Dfb)", in at least the major fraction of its area. For the Lena catchment (Fig. 5.8), we note for the lead year 1 a generally good representation of the TWS anomalies in the dry years 2003 and 2010, that are represented well in both b1-LR and b1-MR. The baseline0 hindcast, however, does not capture those anomalies, its time-variability instead is rather comparable with the uninitialized projections of both model versions. The wet anomaly in 2007 is also captured by both b1-LR and b1-MR, but with much smaller peak amplitudes. In particular, multi-year changes in TWS potentially related to storage changes in deeper soil layers and near-surface aquifers are not reflected in the MPI-ESM model runs. Since feedbacks from those deeper storage reservoirs to the climate system are generally difficult to quantify, it remains open at this point if this deficit of MPI-ESM is important for the quality of decadal climate predictions. Similar conclusions for lead year 1 are obtained for the Nelson catchment. We note a general increase in area-averaged TWS during the years 2003 to 2005 in both b1-LR and b1-MR, which closely resembles the observed variability but is not present in the uninitialized projections. We see a subsequent decline in TWS during the following years until 2008, which is again quite nicely captured by both baseline1 hindcast ensembles. Dry conditions in 2009, however, are only reproduced by b1-LR, whereas the MR hindcasts rather predict a rapid increase in TWS for that year. Focusing on longer

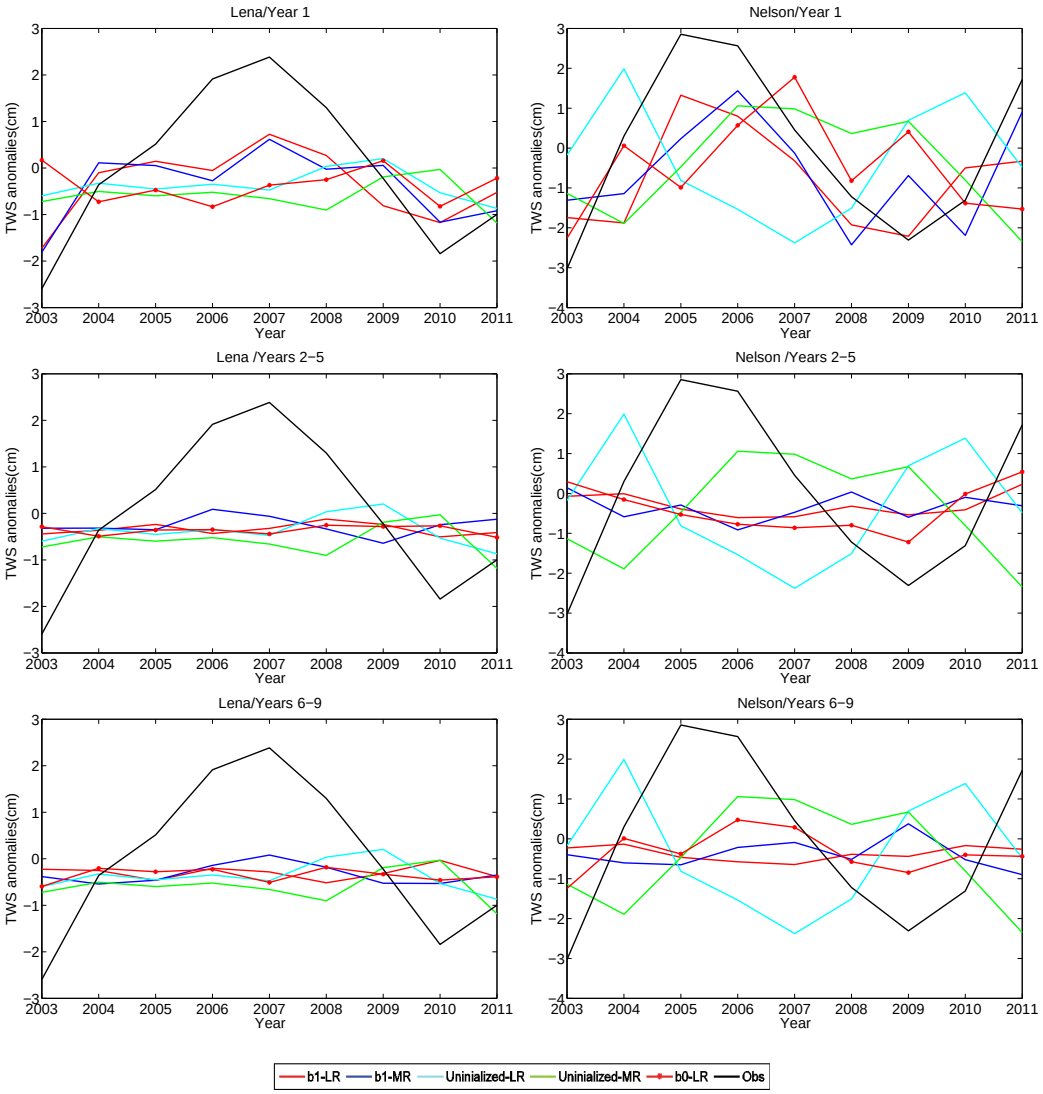

Figure 5.8: Time series of yearly averages of TWS for the Lena catchment in Siberia (left) and the Nelson catchment in North America (right): GRACE-based observations (black), uninitialized projection at low resolution (cyan), uninitialized projection at medium resolution (green), and decadal hindcast experiments b0-LR (red∗), b1-LR (red), and b1-MR (blue).

lead times of 2 - 5 or 6 - 9 years, however, we note substantially smaller differences between hindcasts and projections for both model versions in both catchments considered. For those lead years, the ensemble mean of the hindcasts is more frequently close to a zero-anomaly forecast, supporting our conclusions drawn from Figs. 5.5 and 5.6.

5.2.6 Discussion and conclusion

One decade of terrestrial water storage estimates from GRACE are applied for the validation of three different decadal hindcast experiments with the MPI-ESM coupled climate model. Mean squared skill scores of annual averages of the ensemble means indicate positive skill for the initialized hindcasts in particular for lead year 1 in moderate to high latitudes of the Northern Hemisphere when compared to the uninitialized simulations or their climatologies. In addition, skills gradually increase when moving from b0-LR to b1-LR, and also less pronounced from b1-LR to b1-MR, indicating that changes in the initialization and increased resolution implemented in the different experiments indeed lead to more skillful initialized hindcasts than in the earlier experiments.

For lead years greater than 1, however, skill rapidly drops down towards zero, and also the differences between the three experiments are diminished. It appears that the simulated TWS variability is substantially lower for the uninitialized projections (and the later lead years of the initialized runs), indicating that the models tend to stay closer to its climatology instead of simulating substantial deviations from it, which is, however, not uncommon in numerical modeling experiments of the climate system.

The average level of skill in predicting TWS in all three experiments is quite modest, in particular when compared to more traditionally considered validation variables. This might be related to the relatively short period of observations, but since GRACE is still in operation and a follow-on mission is scheduled to launch in 2017 (Flechtner *et al.*, 2014), this situation will gradually improve during the next years. Further, contamination of GRACE results due to spatial leakage and mass transport processes not related to the global water cycle must be taken into account. In this regard, on-going re-processing efforts and refined post-processing approaches will contribute to a more accurate quantification of random noise and systematic biases in the satellite data.

The good capture of the dry and wet anomalies in the first year of the initialized runs in some areas and the improvement caused by the new initialization strategy is encouraging, given that only oceanic and atmospheric initialization is applied. Earlier studies already demonstrated that land surface initialization is able to additionally contribute - via direct feedbacks into the

atmosphere - to the skill of sub-seasonal and seasonal predictions by, for instance, increasing the boreal summer predictability from soil moisture observations (Dirmeyer, 2005; Koster *et al.*, 2010; Douville, 2010). Since also multi-year memory effects of soil moisture have been found to have promising feedbacks on the climate (Seneviratne *et al.*, 2013), it would be straightforward to also consider GRACE-based TWS as an observable to be assimilated into a land surface re-analysis data-set, that might be subsequently used for initialization of future decadal climate prediction experiments.

5.3 Probabilistic skill metric investigation of MPI-ESM decadal hindcasts with TWS from GRACE

The probabilistic skill of the hindcast experiments is also assessed to quantify the range of possibilities, through the continuous ranked probability skill score (CRPSS). It serves to answer the research question of whether the prediction model's ensemble spread is an appropriate representation of forecast uncertainty on average. The CRPSS is based on the continuous ranked probability score, defined by Eq. 5.8. Since the conditional bias could dominate the probabilistic error in this metric, the H_j here has also been corrected for the conditional bias which is obtained by:

$$\hat{H}_j = (S_0/S_H) r_{HO} H_j, \tag{5.11}$$

where S_O and S_H are the square roots of the variances of the observations and hindcast ensemble means, respectively; and r_{HO} is the correlation between the observations and ensemble means. The corresponding skill score is then expressed as Eq. 5.8. We take b1-LR as an example, and assess the skill score of the average ensemble spread against the standard errors of the mean prediction, the average ensemble spread against the climatological distribution and also the standard error of hindcasts against the climatological distribution (Fig. 5.9). The average ensemble spread is measured as the average standard deviation about the ensemble mean from each ensemble member, and the standard error is calculated as the standard deviation of the residuals from the least squares fit between the observations and the ensemble mean hindcasts.

The application of standard error and average ensemble spread exhibits improved probabilistic skill compared to the climatological distribution (Fig. 5.9, bottom). Goddard *et al.* (2013) indicates similar statements for temperature and precipitation. The CRPSS that tests the uncertainty from the average ensemble spread against the uncertainty from standard error,

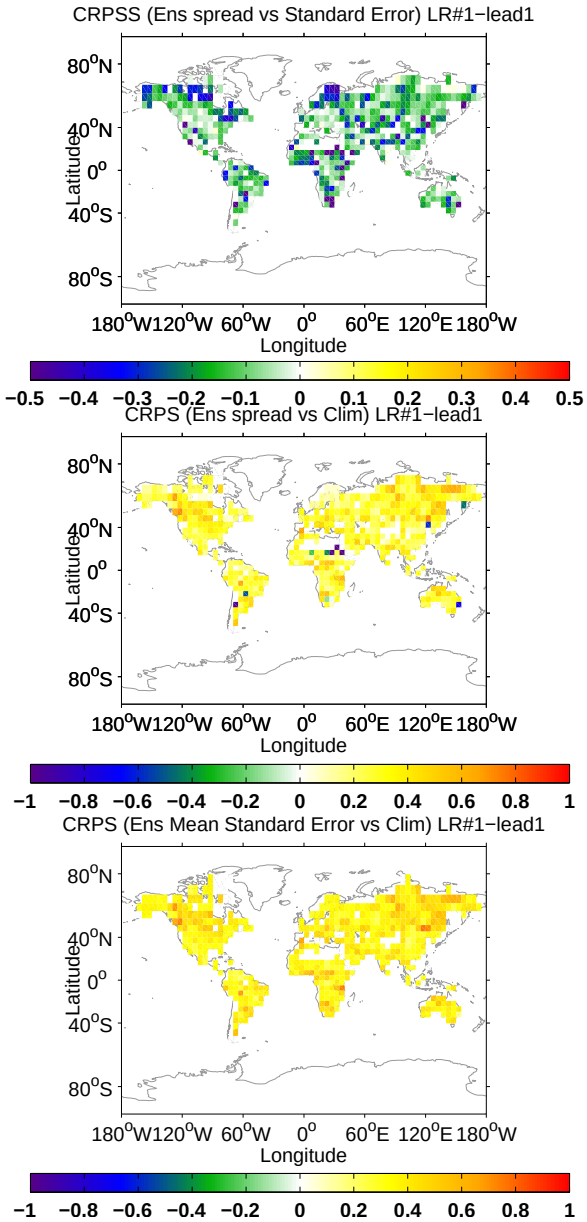

Figure 5.9: Continuous ranked probability skill score (CRPSS) between the b1-LR hindcast experiments with uncertainty of the average ensemble spread against the standard error of the ensemble mean (top); CRPSS with uncertainty given by the average ensemble spread against the climatological distribution (middle), and CRPSS comparing the same hindcasts with uncertainty given by standard error against the climatological distribution (bottom) for lead year 1.

however, shows negative values in most areas (Fig. 5.9, top). This demonstrates that the use of ensemble spread for each individual forecast is less reliable and not suitable as reference for the forecast uncertainty. Considering that the number of ensemble members applied here is still quite small, the ensemble spread can hardly represent the true range of possibilities for the hindcast experiments.

Chapter 6

Conclusions and outlook

In this thesis, TWS variations estimated from GRACE gravity field data have been investigated for hydrometeorological applications, firstly to validate global hydrological models, and secondly to assess the skill performances of the different sets of decadal hindcasts.

Through a closed-loop environment by means of simulated GRACE-type gravity field time-series based on realistic orbits and instrument error assumptions as well as background error assumptions out of the updated ESA Earth System Model, different post-processing strategies are compared and investigated. It is found that the application of DDK2 filter approximately corresponding to a Gaussian radius of 340 km performs slightly better with smaller RMS differences (RMSD) between basin-average H and the GRACE-like TWS variations than the other two versions of this non-isotropic filter (DDK1 and DDK3). The scaling factors to remedy the signal alteration caused by filtering are largely affected by numerical model uncertainties and its contribution to the total TWS error cannot be neglected. The median scaling factors calculated from a small ensemble of multiple hydrological models make them less affected by deficiencies in a certain model. There is, however, still a large gap with the results rescaled with the true hydrology signal (H) in terms of the wRMS of the differences, indicating that further research efforts are needed to narrow this gap, which will significantly reduce the TWS uncertainties. The time variable scaling factors for each month over the whole five years do not yield smaller RMSD values than the constant ones, due to their high sensitivity to the model uncertainties. The globally gridded error estimates from individual contributions of measurement errors, leakage errors, and re-scaling errors exhibit consistent results with the RMSD between the GRACE-like TWS and the "truth" (H, hydrological component of the updated ESA Earth System Model). The TWS variations and uncertainties are then estimated from the real GRACE gravity data (GFZ RL05a) by applying the suggested post-processing strategy from the simulation test. The filtered-only TWS varia-

tions are consistent with the ones from the Tellus website, with the RMSD between the two generally below the estimated TWS error level. Some larger differences are found at certain basins between the rescaled TWS time-series from different models, indicating again the importance of model-based information required to account for spatial leakage. The basin-averaged TWS errors from gridded error estimates show largely consistent results for most basins when compared to estimates that are directly derived from the spherical harmonics representation, which could be used as an upper bound for the GRACE uncertainties.

With almost 14 years of GRACE data available, the TWS variations and uncertainties estimated from this state-of-the-art post-processing strategy are then applied to validate four different global hydrological models. The four statistic metrics investigation indicates that for the 31 largest globally distributed basins, model runs agree with GRACE to a very limited degree only in terms of TWS, with large spreads among the models themselves. Although a common atmospheric forcing data-set has been applied to all hydrological models, potential evapotranspiration (PET) calculated from the models are different due to the various approaches applied, which in a way affect the actual evapotranspiration (AET). Although some large differences of AET are seen, their relationship with the subsequently simulated TWS variations is complicated and in many cases the effects are damped. In addition, the performance of the models varies in different areas as exemplarily shown in the TWS time series comparison in two specific regions: the snow-dominated catchments and the dry catchments. The large model discrepancies are not entirely related to uncertainties in meteorologic input, but instead to the model structure and parametrization, and in particular to the representation of individual storage components with different spatial characteristics in each of the models. The different statistic performances of the models at different basins shows that GRACE-based TWS is quite sensitive to very different model physics in individual basins, which offers helpful insight to modellers for the future improvement of large-scale numerical models of the global terrestrial water cycle. There are still some large differences between TWS from GRACE and models at certain basins (eg. Yukon) that cannot be well explained. With the development of the GRACE gravity solutions and improved model simulations, the reasons for these gaps will be further investigated.

One decade of the TWS estimates from GRACE are applied also for the validation of three different decadal hindcast experiments with the MPI-ESM coupled climate model. Mean squared skill scores of annual averages of the ensemble means indicate positive skill for the initialized hindcasts in particular for lead year 1 in moderate to high latitudes of the Northern Hemisphere when compared to the uninitialized simulations or their climatologies. In addition, skills gradually increase when moving from baseline0 at low resolution (b0-LR) to baseline1 at low resolution (b1-LR), and also less pronounced from b1-LR to baseline1

at medium resolution (b1-MR), indicating that changes in the initialization and increased resolution implemented in the different experiments indeed lead to more skillful initialized hindcasts than in the earlier experiments. For lead years greater than 1, skill rapidly drops down towards zero, and also the differences between the three experiments are diminished. The short time period of GRACE observations still leads to less significant skill score in many areas, which could be improved in future since GRACE is still in operation and the GRACE follow-on mission is scheduled to launch in 2018.

Both applications exhibit high requirement for good quality GRACE-based TWS, where ongoing re-processing efforts and refined post-processing approaches to a more accurate quantification of random noise and systematic biases in the satellite data are needed. The ITSG (Institute of Theoretical Geodesy and Satellite Geodesy) solutions (latest version ITSG-Grace 2016) at Graz University of Technology using variational equations approach is demonstrated to provide better quality GRACE gravity data with reduced noises. Besides, the GRACE-FO is going to be launched soon in 2018, and it not only extends the time series of the gravity field from GRACE, but also uses slightly improved MWI-based gravity field models. There will also be a new laser ranging interferometer (LRI) on-board which provides more precise range observations and could lead to accuracy improved gravity field (Flechtner *et al.*, 2016). There are also several proposals for future gravity missions. For instance, a combination of two pair satellites that fly in different orbits would increase the sampling frequency of the mission and bear higher East-West sensitivity to the gravity field (Wiese *et al.*, 2012). This would, in one way, reduce the level of errors in the gravity solution, in another, require less strong smoothing or filtering which remove the geophysical signals and also decrease the spatial resolution. With the improvement of the accuracy of the gravity field, different post-processing procedures would be needed to pose least leakage and signal reduction. It is then interesting to investigate what more can be learned of the geophysical process on land and whether the targeted science and user needs for observing global water mass transport to understand global change and benefit society found by an expert assessment initiative of the International Union of Geodesy and Geophysics (IUGG) (Pail *et al.*, 2015) would be satisfied.

The thesis mainly focuses on the seasonal variations where the secular trend is removed and the high-frequency signals are not considered. The comparison of the trend of GRACE-like TWS and the "truth" (H) exhibits some large differences even when the same GIA and ice melting signals are removed (not shown in this thesis). Although the trend of TWS is contaminated by the poorly captured trend of ice and solid earth, it is also partly related to the fact that the rescaling factors from hydrological models to compensate for leakage and reduction effects are not suitable beyond the seasonal timescales, as current hydrological models perform poorly to capture those long term and inter-annual signals due to missing

processes and storage parametrization. As shown in areas with intensive irrigation and glacier melting, large variations are found in the re-scaling factors, indicating large uncertainties (Zhang *et al.*, 2016). Within the closed-loop simulation environment as introduced in Section 3.1, where the secular trend of signals from TWS, ice and solid Earth from the source model is already known, it would be interesting to investigate on the effect of a set of two scaling factors for seasonal and long-term variations separately.

The high-frequency signals in GRACE-based TWS are generally not paid attention to, as they are assumed as noise. It is found, however, meaningful hydrological information also exists in the sub-seasonal signals which have high correlation with the daily precipitation through an adequate averaging filter (Humphrey *et al.*, 2016). Daily GRACE gravity field model series also demonstrate their ability to capture the high-frequency non-tidal mass variations, such as the extra-tropical sea-level variability in the Southern Ocean (Zenner *et al.*, 2014). Although the short-term TWS variability is not well simulated by current hydrological models, the reconstructed high-frequency TWS signals from the daily precipitation and the daily GRACE gravity field products can serve as alternative ways for GRACE de-aliasing (Eicker & Springer, 2016). The reconstructed high-frequency signals provide a new perspective to assess GRACE uncertainties and may help to explain the differences between TWS from GRACE and global numerical models, which will be further investigated in future.

Appendix A

Filters

A.1 Swenson-Wahr filter

Swenson & Wahr (2006) indicated that the stripes in the surface mass variability estimates are related to correlations in the spectral domain for even and odd degrees, separately. First, smoothed Stokes coefficients for a particular order (m) are obtained by fitting a quadratic polynomial in a moving window centered at a certain degree l which is done for even and odd separately and then such correlated-error coefficients are subtracted to remove the stripes. The smoothed coefficients can be expressed as a combination of the un-filtered coefficients:

$$C^s_{lm} = \sum_{\substack{n=l-w/2 \\ n:even\ or\ odd}}^{l+w/2} \Lambda_{lnm} C_{nm}, \tag{A.1}$$

and the filter is defined as:

$$\Lambda_{lnm} = \sum_{i=0}^{p} \sum_{j=0}^{p} L^{-1}_{ij} n^j l^i, \tag{A.2}$$

where p is the order of the polynomial, and in a quadratic case, $p = 2$. w is the window width, and only the terms with the same parity as l within w are summed. As indicated by Swenson & Wahr (2006), the correlations occur at approximately $m = 8$ and then are present in all high degrees. The filter is then only applied to the orders starting from 8. The window width w decreases with increasing orders and can be express in the following equation according to Duan *et al.* (2009):

$$w = max(Ae^{-m/K} + 1, 5) \tag{A.3}$$

A and K are empirically chosen as 30 and 10 based on a trial-and-error procedure. The small window width for higher orders is chosen to apply a stronger filter in higher orders which are much noisier. When the w becomes smaller than 5, the window width is defined as 5 to assure the minimum number of degrees for a polynomial least square fit. The choice of the parameters p, w and m could also be different depend on the applied GRACE gravity data and specific applications. Chambers (2006) has chosen $m = 11$ for the CSR RL04 data, and applied fifth order polynomial to fit all the spherical harmonics of the same order for even and odd degrees seperately. As the purpose of Chambers (2006) is to apply the method on the ocean, which has much smaller signals compared with land area, a stronger filter has been applied. Chen *et al.* (2007) kept the spherical harmonics of orders 5 and below as they were and calculated the correlations by applying a third order polynomial to fit the spherical harmonics of the same order. Since he was interested in the 2004 Sumatra-Andaman earthquake induced mass changes, the comparably weaker filter was applied to exibit more detailed information around the equatorial region. This also caused the stripes to be less reduced.

A.2 DDK filter

The concept of DDK filter proposed by (Kusche, 2007) is in some way comparable to the Swenson & Wahr (2002) filter, both of which have applied a priori signal and error covariance matrices. For Swenson & Wahr (2002) filter, the error covariance matrix is obtained by an empirical inspection of the coefficients, while Kusche (2007) imitates the regularization of GRACE data processing, and derives the error covariance matrix synthetically from one set of GRACE monthly normal equations. The filtered Stokes coefficients can be expressed as:

$$\hat{x}_\gamma = W_\gamma \hat{x}, \tag{A.4}$$

where W_γ is the matrix of decorrelation coefficients. Unlike Gaussian filtering, which is only degree dependent, the weights in the W matrix of DDK and the Swenson & Wahr (2002) filter depend on both the degree and order of th SH coefficients:

$$\hat{x}_{lmq;\gamma} = w_{lmq,\gamma} \hat{x}_{lmq}, \tag{A.5}$$

and the x_{lmq} is expressed as:

$$x_{lmq} = \{ \begin{array}{cc} C_{lm} & for \quad q = 1 \\ S_{lm} & \quad\quad q = 2 \end{array} . \tag{A.6}$$

Table A.1: The corresponding Gaussian radius and parameter a and p for the decorrelation filters from DDK1 to DDK8 (Kusche *et al.*, 2009).

Decorrelation filter	Corresponding Gaussian radius (km)	Parameter a and p
DDK1	530	$a = 1 \times 10^{14}$, $p = 4$
DDK2	340	$a = 1 \times 10^{13}$, $p = 4$
DDK3	240	$a = 1 \times 10^{12}$, $p = 4$
DDK4	220	$a = 5 \times 10^{11}$, $p = 4$
DDK5	160	$a = 1 \times 10^{11}$, $p = 4$
DDK6	140	$a = 5 \times 50^{10}$, $p = 4$
DDK7	120	$a = 1 \times 10^{10}$, $p = 4$
DDK8	100	$a = 5 \times 50^{9}$, $p = 4$

The W for DDK filter is obtained by a least squares adjustment. The Stokes coefficients, denoted as the parameters x, are obtained from:

$$x = N^{-1}b, \tag{A.7}$$

where N is the normal equation matirx and b is the right hand side vector. A regularized estimate for the parameters is given by:

$$x_\gamma(a) = (N + aM)^{-1}b = (N + aM)^{-1}Nx = W_\gamma(a)x, \tag{A.8}$$

where N^{-1} is the GRACE error covariance matrix and M^{-1} the a priori signal covariance matrix. By tuning the regularization parameter a, it is possible to downweight the signal covariance and change the strength of smoothing.

There are eight different filter matrices $W_\gamma(a)$, for DDK1 to DDK8 available, where DDK1 implies the most and DDK8 the least smoothing. Their corresponding choice of the regularization parameter a is given in Table A.1. The filter effect depends on latitude and is stronger in the East-West than in North-South direction, so that it adapts well to the GRACE error structure. The decorrelation effect is mainly due to the negative side lobes.

References

Balmaseda, M. A., Mogensen, K. & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4, *Quart. J. Roy. Meteor.*, **139**, 1132–1161, doi:10.1002/qj.2063, URL `http://dx.doi.org/10.1002/qj.2063`.

Bergmann, I. & Dobslaw, H. (2012). Short-term transport variability of the antarctic circumpolar current from satellite gravity observations, *J. Geophys. Res.*, **117**, doi:10.1029/2012JC007872.

Bergmann-Wolf, I. (2015). *Oceanographic applications of GRACE gravity data on global and regional scales*, Ph.D. thesis, Berlin free University.

Bergmann-Wolf, I., Dill, R., Forootan, E., Klemann, V., Kusche, J., Sasgen, I. & Dobslaw, H. (2014a). *Updating ESA's Earth System Model for Gravity Mission Simulation Studies: 2. Comparison with the Original Model*, *GFZ Scientific Technical Report*, GFZ, Helmholtz-Zentrum, Potsdam, Potsdam.

Bergmann-Wolf, I., Zhang, L. & Dobslaw, H. (2014b). Global eustatic sea-level variations for the approximation of geocenter motion from GRACE, *J. Geod. Sci.*, **4**, 37–48, doi:10.2478/jogs-2014-0006.

Biancale, R. & Bode, A. (2006). *Mean annual and seasonal atmospheric tide models based on 3-hourly and 6-hourly ECMWF surface pressure data* , *Scientific Technical Report STR06/01*, GFZ, Helmholtz-Zentrum, Potsdam, doi:10.2312/GFZ.b103-06011.

Bierkens, M. F. P. & van den Hurk, B. J. J. M. (2007). Groundwater convergence as a possible mechanism for multi-year persistence in rainfall, *Geophys. Res. Lett.*, **34**, doi:10.1029/2006GL028396, URL `http://dx.doi.org/10.1029/2006GL028396`.

Blewitt, G., Lavallée, D., Clarke, P. & Nurutdinov, K. (2001). A new global mode of earth deformation: Seasonal cycle detected, *Science*, **294**, 2342–2345, doi:10.1126/science.1065328, URL `http://science.sciencemag.org/content/294/5550/2342`.

Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M. & Gayler, V. (2009). Global biogeophysical interactions between forest and climate, *Geophys. Res. Lett.*, **36**, doi:10.1029/2009GL037543.

Cazenave, A. & Chen, J. (2010). Time-variable gravity from space and present-day mass redistribution in the earth system, *Earth and Planetary Science Letters*, **298**, 263 – 274, doi:http://dx.doi.org/10.1016/j.epsl.2010.07.035, URL `http://www.sciencedirect.com/science/article/pii/S0012821X10004735`.

Chambers, D. P. (2006). Evaluation of new GRACE time-variable gravity data over the ocean, *Geophysical Research Letters*, **33**, n/a–n/a, doi:10.1029/2006GL027296, URL `http://dx.doi.org/10.1029/2006GL027296`, l17603.

Chambers, D. P. & Bonin, J. A. (2012). Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean, *Ocean Sci.*, **8**, 859–868, doi:10.5194/os-8-859-2012.

Chambers, D. P., Wahr, J. & Nerem, R. S. (2004). Preliminary observations of global ocean mass variations with GRACE, *Geophysical Research Letters*, **31**, n/a–n/a, doi:10.1029/2004GL020461, URL `http://dx.doi.org/10.1029/2004GL020461`, l13310.

Chambers, D. P., Wahr, J., Tamisiea, M. E. & Nerem, R. S. (2010). Ocean mass from GRACE and glacial isostatic adjustment, *J. Geophys. Res., B*, **115**, doi:10.1029/2010JB007530, b11415.

Chen, J., Wilson, C., Famiglietti, J. & Rodell, M. (2007). Attenuation effect on seasonal basin-scale water storage changes from GRACE time-variable gravity, *J. Geodesy*, **81**, 237–245, doi:10.1007/s00190-006-0104-2.

Chen, Q., Shen, Y., Chen, W., Zhang, X. & Hsu, H. (2016). An improved GRACE monthly gravity field solution by modeling the non-conservative acceleration and attitude observation errors, *Journal of Geodesy*, **90**, 503–523, doi:10.1007/s00190-016-0889-6, URL `http://dx.doi.org/10.1007/s00190-016-0889-6`.

Cheng, M., Ries, J. C. & Tapley, B. D. (2011). Variations of the Earth's figure axis from satellite laser ranging and GRACE, *J. Geophys. Res.*, **116**, 1409–+, doi:10.1029/2010JB000850.

Colombo, O. L. (1984). The global mapping of gravity with two satellites, *Netherlands Geodetic Commission*, **7**, publications on Geodesy.

Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G. & Neumayer, K. (2012). *GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005* , *Scientific technical report-data*, GFZ, Helmholtz-Zentrum, Potsdam, Potsdam, doi:10.2312/GFZ.b103-1202-25.

Davis, J. L., Elósegui, P., Mitrovica, J. X. & Tamisiea, M. E. (2004). Climate-driven deformation of the solid earth from grace and gps, *Geophysical Research Letters*, **31**, n/a–n/a, doi:10.1029/2004GL021435, URL `http://dx.doi.org/10.1029/2004GL021435`, l24605.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. & Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quart. J. Roy. Meteor.*, **137**, 553–597, doi:10.1002/qj.828, URL `http://dx.doi.org/10.1002/qj.828`.

Derber, J. C., Parrish, D. F., & Lord, S. J. (1991). The New Global Operational Analysis System at the National Meteorological Center, *Wea. Forecasting*, **6**, 538–547, doi:10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2.

Dill, R. (2008). *Hydrological model LSDM for operational Earth rotation and gravity field variations*, *GFZ Scientific Technical Report-STR08/09*, GFZ, Helmholtz-Zentrum, Potsdam, Potsdam.

Dill, R. & Dobslaw, H. (2010). Short-term polar motion forecasts from earth system modeling data, *J. Geodesy*, **84**, 529–536, doi:10.1007/s00190-010-0391-5.

Dill, R. & Dobslaw, H. (2013). Numerical simulations of global-scale high-resolution hydrological crustal deformations, *J. Geophys. Res.: B*, **118**, 5008–5017, doi:10.1002/jgrb.50353.

Dirmeyer, P. A. (2005). The land surface contribution to the potential predictability of boreal summer season climate, *J. Hydrometeor.*, **6**, 618–632, doi:http://dx.doi.org/10.1175/JHM444.1.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., & Hanasaki, N. (2006). GSWP-2: Multimodel analysis and implications for our perception of the land surface, *Bull. Amer. Meteor. Soc.*, **87**, 1381–1397, doi:110.1175/BAMS-87-10-1381.

Dirmeyer, P. A., Schlosser, C. A. & Brubaker, K. L. (2009). Precipitation, recycling, and land memory: An integrated analysis, *Journal of Hydrometeorology*, **10**, 278–288, doi:10.1175/2008JHM1016.1, URL `http://dx.doi.org/10.1175/2008JHM1016.1`.

Ditmar, P. (2015). How to properly convert stokes coefficients into mass anomalies?, EGU General Assembly, Vienna, Austria.

Dobslaw, H., Bergmann-Wolf, I., Dill, R., Forootan, E., Klemann, V., Kusche, J. & Sasgen, I. (2015). The updated ESA Earth System Model for future gravity mission simulation studies, *Journal of Geodesy*, **89**, 505–513, doi:10.1007/s00190-014-0787-8, URL `http://dx.doi.org/10.1007/s00190-014-0787-8`.

Dobslaw, H., Bergmann-Wolf, I., Forootan, E., Dahle, C., Mayer-Gürr, T., Kusche, J. & Flechtner, F. (2016). Modeling of present-day atmosphere and ocean non-tidal de-aliasing errors for future gravity mission simulations, *Journal of Geodesy*, **90**, 423–436, doi:10.1007/s00190-015-0884-3, URL `http://dx.doi.org/10.1007/s00190-015-0884-3`.

Dobslaw, H., Dill, R., Grötzsch, A., Brzeziński, A. & Thomas, M. (2010). Seasonal polar motion excitation from numerical models of atmosphere, ocean, and continental hydrosphere, *J. Geophys. Res.: B*, **115**, doi:10.1029/2009JB007127.

Dobslaw, H., Flechtner, F., Bergmann-Wolf, I., Dahle, C., Dill, R., Esselborn, S., Sasgen, I. & Thomas, M. (2013). Simulating high-frequency atmosphere-ocean mass variability for dealiasing of satellite gravity observations: AOD1B RL05, *J. Geophys. Res.: Oceans*, doi:10.1002/jgrc.20271.

Dobslaw, H. & Thomas, M. (2007). Simulation and observation of global ocean mass anomalies, *Journal of Geophysical Research: Oceans*, **112**, n/a–n/a, doi:10.1029/2006JC004035, URL `http://dx.doi.org/10.1029/2006JC004035`, c05040.

Döll, P., Kaspar, F. & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol*, **270**, 105–134, doi:10.1016/S0022-1694(02)00283-4.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A. & Jackson, T. (2011). The international soil moisture network: a data hosting facility for global in situ soil moisture measurements, *Hydrology and Earth System Sciences*, **15**, 1675–1698, doi:10.5194/hess-15-1675-2011, URL `http://www.hydrol-earth-syst-sci.net/15/1675/2011/`.

Douville, H. (2010). Relative contribution of soil moisture and snow mass to seasonal climate predictability: a pilot study, *Clim. Dyn.*, **34**, 797–818, doi:10.1007/s00382-008-0508-1, URL `http://dx.doi.org/10.1007/s00382-008-0508-1`.

Duan, X. J., Guo, J. Y., Shum, C. K. & van der Wal, W. (2009). On the postprocessing removal of correlated errors in GRACE temporal gravity field solutions, *Journal of Geodesy*, **83**, 1095–1106, doi:10.1007/s00190-009-0327-0, URL `http://dx.doi.org/10.1007/s00190-009-0327-0`.

Eanes, R. (2000). *SLR solutions from the University of Texas Center for Space Research, Geocenter from TOPEX SLR/DORIS, 1992-2000*, *Tech. rep.*, Pasadena, Calif., URL `http:///sbgg.jpl.nasa.gov/dataset.html`, iERS Spec. Bur. for Gravity/Geocent.

Eicker, A. (2008). *Gravity field refinement by radial basis functions from in-situ satellite data*, Ph.D. thesis, Univ. Bonn.

Eicker, A., Schumacher, M., Kusche, J., Döll, P. & Schmied, H. M. (2014). Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results , *Surveys in Geophysics*, **35**, 1285–1309, doi:10.1007/s10712-014-9309-8.

Eicker, A. & Springer, A. (2016). Monthly and sub-monthly hydrological variability: in-orbit validation by grace level 1b observations, *Journal of Geodesy*, **90**, 573–584, doi:10.1007/s00190-016-0895-8, URL `http://dx.doi.org/10.1007/s00190-016-0895-8`.

Elsaka, B., Raimondo, J.-C., Brieden, P., Reubelt, T., Kusche, J., Flechtner, F., Iran Pour, S., Sneeuw, N. & Müller, J. (2014). Comparing seven candidate mission configurations for temporal gravity field retrieval through full-scale numerical simulation, *Journal of Geodesy*, **88**, 31–43, doi:10.1007/s00190-013-0665-9, URL `http://dx.doi.org/10.1007/s00190-013-0665-9`.

Ettema, J., van den Broeke, M. R., van Meijgaard, E., van de Berg, W. J., Bamber, J. L., Box, J. E. & Bales, R. C. (2009). Higher surface mass balance of the greenland ice sheet revealed by high-resolution climate modeling, *Geophysical Research Letters*, **36**, n/a–n/a, doi:10.1029/2009GL038110, URL `http://dx.doi.org/10.1029/2009GL038110`, l12501.

Famiglietti, J. S. (2014). The global groundwater crisis, *Nature Climate Change*, **4**, 945–948, doi:10.1038/nclimate2425.

Famiglietti, J. S. & Rodell, M. (2013). Water in the balance, *Science*, **340**, 1300–1301, doi:10.1126/science.1236460.

Farrell, W. E. (1972). Deformation of the Earth by surface loads, *Reviews of Geophysics*, **10**, 761–797, doi:10.1029/RG010i003p00761, URL `http://dx.doi.org/10.1029/RG010i003p00761`.

Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H.-T. & Xia, J. (2013). Evaluation of groundwater depletion in north china using the gravity recovery and climate experiment (grace) data and ground-based measurements, *Water Resources Research*, **49**, 2110–2118, doi:10.1002/wrcr.20192, URL `http://dx.doi.org/10.1002/wrcr.20192`.

Fischer, E. M. & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes, *Nature Climate Change*, **5**, 560–564, doi:10.1038/nclimate2617.

Flechtner, F. & Dobslaw, H. (2013). *AOD1B product description document for product Release 05*, *Scientific Technical Report* , GFZ, Helmholtz-Zentrum, Potsdam.

Flechtner, F., Neumayer, K.-H., Dahle, C., Dobslaw, H., Fagiolini, E., Raimondo, J.-C. & Güntner, A. (2016). What can be expected from the GRACE-FO Laser Ranging Interferometer for Earth science applications?, *Surveys in Geophysics*, **37**, 453–470, doi:10.1007/s10712-015-9338-y, URL `http://dx.doi.org/10.1007/s10712-015-9338-y`.

Flechtner, F., Watkins, M., Morton, P. & Webb, F. (2014). Status of the GRACE Follow-on Mission, in: *Proceedings of the International Association of Geodesy Symposia Gravity, Geoid and Height System (GGHS2012)*, vol. IAGS-D-12-00141, Venice, Italy.

Förste, C., Schmidt, R., Stubenvoll, R., Flechtner, F., Meyer, U., König, R., Neumayer, H., Biancale, R., Lemoine, J.-M., Bruinsma, S., Loyer, S., Barthelmes, F. & Esselborn, S. (2008). The GeoForschungsZentrum Potsdam/Groupe de Recherche de Gèodésie Spatiale satellite-only and combined gravity field models: EIGEN-GL04S1 and EIGEN-GL04C, *Journal of Geodesy*, **82**, 331–346, doi:10.1007/s00190-007-0183-8, URL `http://dx.doi.org/10.1007/s00190-007-0183-8`.

Frappart, F., Seoane, L. & Ramillien, G. (2013). Validation of GRACE-derived terrestrial water storage from a regional approach over South America, *Remote Sensing of Environment*, **137**, 69–83, doi:http://dx.doi.org/10.1016/j.rse.2013.06.008.

Fritsche, M., Dietrich, R., Rülke, A., Rothacher, M. & Steigenberger, P. (2010). Low-degree earth deformation from reprocessed GPS observations, *GPS Solutions*, **14**, 165–175, doi:10.1007/s10291-009-0130-7, URL `http://dx.doi.org/10.1007/s10291-009-0130-7`.

García-Serrano, J. & Doblas-Reyes, F. J. (2012). On the assessment of near-surface global temperature and north atlantic multi-decadal variability in the ensembles decadal hindcast, *Climate Dynamics*, **39**, 2025–2040, doi:10.1007/s00382-012-1413-1, URL `http://dx.doi.org/10.1007/s00382-012-1413-1`.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J.,

Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J. & Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *Journal of Advances in Modeling Earth Systems*, **5**, 572–597, doi:10.1002/jame.20038, URL `http://dx.doi.org/10.1002/jame.20038`.

Goddard, L., Barnston, A. G. & Mason, S. J. (2003). Evaluation of the IRI'S "Net Assessment" Seasonal Climate Forecasts: 1997–2001, *Bulletin of the American Meteorological Society*, **84**, 1761–1781, doi:10.1175/BAMS-84-12-1761, URL `http://dx.doi.org/10.1175/BAMS-84-12-1761`.

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., Mason, S., Kirtman, B., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C., Stephenson, D., Meehl, G., Stockdale, T., Burgman, R., Greene, A., Kushnir, Y., Newman, M., Carton, J., Fukumori, I. & Delworth, T. (2013). A verification framework for interannual-to-decadal predictions experiments, *Clim. Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2, URL `http://dx.doi.org/10.1007/s00382-012-1481-2`.

Gruber, T., Bamber, J. L., Bierkens, M. F. P., Dobslaw, H., Murböck, M., Thomas, M., van Beek, L. P. H., van Dam, T., Vermeersen, L. L. A. & Visser, P. N. A. M. (2011). Simulation of the time-variable gravity field by means of coupled geophysical models, *Earth System Science Data*, **3**, 19–35, doi:10.5194/essd-3-19-2011, URL `http://www.earth-syst-sci-data.net/3/19/2011/`.

Gudmundsson, L., Wagener, T., Tallaksen, L. M. & Engeland, K. (2012). Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, *Water Resour. Res.*, **48**, doi:10.1029/2011WR010911, W11504.

Gunkel, A. & Lange, J. (2011). New insights into the natural variability of water resources in the lower Jordan river Basin, *Water Resour. Manag.*, **26**, 963–980, doi:10.1007/s11269-011-9903-1.

Güntner, A. (2008). Improvement of Global Hydrological Models Using GRACE Data, *Surv Geophys*, **29**, 375–397.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K. & Merz, B. (2007). A global analysis of temporal and spatial variations in continental water storage, *Water Resour. Res.*, **43**, doi:10.1029/2006WR005247, W05416.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voss, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., Yehm, P., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R. & Heinke, J. (2011). Multimodel estimate of the global terrestrial water balance: Setup and first results, *J. Hydrometeor*, **12**, 869–884, doi:10.1175/2011JHM1324.1.

Hagemann, S. & Gates, L. (2001). Validation of the hydrological cycle of ECMWF and NCEP reanalyses using the MPI hydrological discharge model, *J. Geophys. Res.*, **106**, 1503–1510, doi:10.1029/2000JD900568.

Hagemann, S. & Gates, L. (2003). Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations, *Clim. Dyn.*, **21**, 349–359, doi:10.1007/s00382-003-0349-x.

Hagemann, S., Loew, A. & Andersson, A. (2013). Combined evaluation of MPI-ESM land surface water and energy fluxes, *J. Adv. Model. Earth Syst.*, **5**, 259–286, doi:10.1029/2012MS000173.

Han, S.-C., Shum, C. K., Jekeli, C., Kuo, C.-Y., Wilson, C. & Seo, K.-W. (2005). Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement, *Geophysical Journal International*, **163**, 18–25, doi:10.1111/j.1365-246X.2005.02756.x, URL `http://dx.doi.org/10.1111/j.1365-246X.2005.02756.x`.

Hedin, A. E. (1987). MSIS-86 thermospheric model, *Journal of Geophysical Research: Space Physics*, **92**, 4649–4662, doi:10.1029/JA092iA05p04649, URL `http://dx.doi.org/10.1029/JA092iA05p04649`.

Heiskanen, W. A. & Moritz, H. (1967). *Physical Geodesy*, Freeman.

Hirschi, M., Seneviratne, S. & Schär, C. (2006). Seasonal variations in terrestrial water storage for major midlatitude river basins, *J. Hydrometeor*, **7**, 39–60, doi:10.1175/JHM480.1.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., uchwitz, M. B., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendael, M., & Wagner, W. (2013). The ESA climate change initiative: satellite data records for essential climate variables, *Bull. Amer. Meteor. Soc.*, **94**, 1541–1552, doi:10.1175/BAMS-D-11-00254.1.

Houborg, R., Rodell, M., Li, B., Reichle, R. & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, *Water Resour. Res.*, **48**, doi:10.1029/2011WR011291, W07525.

Humphrey, V., Gudmundsson, L. & Seneviratne, S. I. (2016). Assessing global water storage variability from GRACE: Trends, seasonal cycle, subseasonal anomalies and extremes, *Surveys in Geophysics*, **37**, 357–395, doi:10.1007/s10712-016-9367-1, URL `http://dx.doi.org/10.1007/s10712-016-9367-1`.

Hunger, M. & Döll, P. (2008). Value of river discharge data for global-scale hydrological modeling, *Hydro. Earth Syst. Sc.*, **12**, 841–861.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H. & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, *Journal of Advances in Modeling Earth Systems*, **5**, 287–315, doi:10.1029/2012MS000178, URL `http://dx.doi.org/10.1029/2012MS000178`.

Jacob, T., Wahr, J., Pfeffer, W. T. & Swenson, S. (2012). Recent contributions of glaciers and ice caps to sea level rise, *Nature*, **481**, 514–518, doi:10.1038/nature10847.

Jekeli, C. (1981). *Alternative Methods to Smooth the Earth's Gravity Field*, Department of Geodetic Science and Surveying: Report, Ohio State University Department of Geodetic Science and Surveying, URL https://books.google.de/books?id=bo8tGwAACAAJ.

Jeong, J.-H., Linderholm, H. W., Woo, S.-H., Folland, C., Kim, B.-M., Kim, S.-J. & Chen, D. (2013). Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season, *Journal of Climate*, **26**, 1956–1972, doi:10.1175/JCLI-D-12-00159.1, URL http://dx.doi.org/10.1175/JCLI-D-12-00159.1.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D. & von Storch, J. S. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, *JAMES*, **5**, 422–446, doi:10.1002/jame.20023, URL http://dx.doi.org/10.1002/jame.20023.

Kadow, C., Illing, S., Kunst, O., Rust, H. W., Pohlmann, H., Müller, W. A. & Cubasch, U. (2015). Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system, *Meteorologische Zeitschrift*, –, doi:10.1127/metz/2015/0639, URL http://dx.doi.org/10.1127/metz/2015/0639.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A. & Reynolds, R. (1996). The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kaula, W. M. (1966). *Theory of satellite geodesy*, Blaisdell Publishing Company.

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L. & Roeckner, E. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, **453**, 84–88, doi:10.1038/nature06921, URL http://dx.doi.org/10.1038/nature06921.

Klees, R., Liu, X., Wittwer, T., Gunter, B. C., Revtova, E. A., Tenzer, R., Ditmar, P., Winsemius, H. C. & Savenije, H. H. G. (2008). A comparison of global and regional grace models for land hydrology, *Surveys in Geophysics*, **29**, 335–359, doi:10.1007/s10712-008-9049-8, URL http://dx.doi.org/10.1007/s10712-008-9049-8.

Klees, R., Zapreeva, E. A., Winsemius, H. C. & Savenije, H. H. G. (2007). The bias in GRACE estimates of continental water storage variations, *Hydrol. Earth Syst. Sc.*, **11**, 1227–1241, doi:10.5194/hess-11-1227-2007, URL http://www.hydrol-earth-syst-sci.net/11/1227/2007/.

Kleidon, A., Renner, M. & Porada, P. (2014). Estimates of the climatological land surface energy and water balance derived from maximum convective power, *Hydrology and Earth System Sciences*, **18**, 2201–2218, doi:10.5194/hess-18-2201-2014, URL http://www.hydrol-earth-syst-sci.net/18/2201/2014/.

Klinger, B., Baur, O. & Mayer-Gürr, T. (2014). GRAIL gravity field recovery based on the short-arc integral equation technique: Simulation studies and first real data results, *Planetary and Space Science*, **91**, 83 – 90, doi:http://dx.doi.org/10.1016/j.pss.2013.12.001, URL http://www.sciencedirect.com/science/article/pii/S0032063313003255.

Klinger, B., Mayer-Gürr, T., Behzadpour, S., Ellmer, M., Kvas, A. & Zehentner, N. (2016). The new ITSG-Grace2016 release, EGU General Assembly, Vienna, Austria, doi:10.13140/RG.2.1.1856.7280.

Knocke, P., Ries, J. & Tapley, B. (1988). Earth radiation pressure effects on satellites, AIAA/AAS astrodynamics specialist conference, Washington DC, USA, p. 577–586.

Köhl, A. (2015). Evaluation of the GECCO2 ocean synthesis: transports of volume, heat and freshwater in the Atlantic, *Quarterly Journal of the Royal Meteorological Society*, **141**, 166–181, doi:10.1002/qj.2347, URL `http://dx.doi.org/10.1002/qj.2347`.

Kopp, T. J. & Kiess, R. B. (1996). The Air Force Global Weather Central cloud analysis model, AMS 15th Conf. on Weather Analysis and Forecasting, Norfolk, VA, p. 220–222.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y. & Yamada, T. (2004a). Regions of strong coupling between soil moisture and precipitation, *Science*, **305**, 1138–1140, doi:10.1126/science.1100217, URL `http://www.sciencemag.org/content/305/5687/1138.abstract`.

Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J.-H., Lawrence, D. M., Lee, W.-S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., Stanelle, T., van den Hurk, B. J. J. M., Vitart, F. & Wood, E. F. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment, *Geophys. Res. Lett.*, **37**, doi:10.1029/2009GL041677, URL `http://dx.doi.org/10.1029/2009GL041677`.

Koster, R. D., Suarez, M. J., Liu, P., Jambor, U., Berg, A., Kistler, M., Reichle, R., Rodell, M. & Famiglietti, J. (2004b). Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *Journal of Hydrometeorology*, **5**, 1049–1063, doi:10.1175/JHM-387.1, URL `http://dx.doi.org/10.1175/JHM-387.1`.

Kottek, M. & Grieser, J. ().

Kruschke, T., Rust, H. W., Kadow, C., Müller, W. A., Pohlmann, H., Leckebusch, G. C. & Ulbrich, U. (2015). Probabilistic evaluation of decadal prediction skill regarding northern hemisphere winter storms, *Meteorologische Zeitschrift*, –, doi:10.1127/metz/2015/0641, URL `http://dx.doi.org/10.1127/metz/2015/0641`.

Kuhlmann, J., Dobslaw, H. & Thomas, M. (2011). Improved modeling of sea level patterns by incorporating self-attraction and loading, *Journal of Geophysical Research: Oceans*, **116**, n/a–n/a, doi:10.1029/2011JC007399, URL `http://dx.doi.org/10.1029/2011JC007399`, c11036.

Kusche, J. (2007). Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J. Geodesy*, **81**, 733–749, doi:10.1007/s00190-007-0143-3.

Kusche, J., Schmidt, R., Petrovic, S. & Rietbroek, R. (2009). Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, *J. Geodesy*, **83**, 903–913, doi:10.1007/s00190-009-0308-3.

Kusche, J. & Schrama, E. J. O. (2005). Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data, *Journal of Geophysical Research: Solid Earth*, **110**, n/a–n/a, doi:10.1029/2004JB003556, URL `http://dx.doi.org/10.1029/2004JB003556`, b09409.

Landerer, F. W. & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, **48**, doi:10.1029/2011WR011453, w04531.

Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O. & Fakes, A. (2009). Basin-scale, integrated observations of the early 21st century multiyear drought in southeast australia, *Water Resources Research*, **45**, n/a–n/a, doi:10.1029/2008WR007333, URL `http://dx.doi.org/10.1029/2008WR007333`, w04408.

Lemoine, F. G., Smith, D. E., Kunz, L., Smith, R., Pavlis, E. C., Pavlis, N. K., Klosko, S. M., Chinn, D. S., Torrence, M. H., Williamson, R. G., Cox, C. M., Rachlin, K. E., Wang, Y. M., Kenyon, S. C., Salman, R., Trimmer, R., Rapp, R. H. & Nerem, R. S. (1997). *The Development of the NASA GSFC and NIMA Joint Geopotential Model*, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 461–469, doi:10.1007/978-3-662-03482-8_62, URL `http://dx.doi.org/10.1007/978-3-662-03482-8_62`.

Lemoine, J. M., Bruinsma, S., Loyer, S., Biancale, R., Marty, J. C., Perosanz, F. & Balmino, G. (2007). Temporal gravity field models inferred from GRACE data, *Advances in Space Research*, **39**, 1620–1629, doi:10.1016/j.asr.2007.03.062, URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-34347209050&partnerID=40&md5=fe09211be54f2e9888ea6f841096a266`.

Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D. & van Dam, T. M. (2012). Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe, *Journal of Hydrology*, **446-447**, 103 – 115, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.04.035, URL `http://www.sciencedirect.com/science/article/pii/S0022169412003228`.

Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres*, **99**, 14,415–14,428, doi:10.1029/94JD00483, URL `http://dx.doi.org/10.1029/94JD00483`.

Llovel, W., Becker, M., Cazenave, A., Crétaux, J.-F. & Ramillien, G. (2010). Global land water storage change from GRACE over 2002-2009; Inference on sea level , *Comptes Rendus Geoscience*, **342**, 179 – 188, doi:http://dx.doi.org/10.1016/j.crte.2009.12.004, URL `http://www.sciencedirect.com/science/article/pii/S1631071310000039`.

Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F. & Syed, T. H. (2010). Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, *Water Resources Research*, **46**, n/a–n/a, doi:10.1029/2009WR007855, URL `http://dx.doi.org/10.1029/2009WR007855`, w05517.

Long, D., Longuevergne, L. & Scanlon, B. R. (2015). Global analysis of approaches for deriving total water storage changes from GRACE satellites, *Water Resour. Res.*, **51**, 2574–2594, doi:10.1002/2014WR016853.

Longuevergne, L., Scanlon, B. R. & Wilson, C. R. (2010). GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, *Water Resour. Res.*, **46**, doi:10.1029/2009WR008564.

Loomis, B. D., Nerem, R. S. & Luthcke, S. B. (2012). Simulation study of a follow-on gravity mission to GRACE, *Journal of Geodesy*, **86**, 319–335, doi:10.1007/s00190-011-0521-8, URL `http://dx.doi.org/10.1007/s00190-011-0521-8`.

Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J. & Camp, J. (2013). Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution, *Journal of Glaciology*, **59**, 613–631, doi:10.3189/2013JoG12J147.

Mahmood, R., Sr., R. A. P., Hubbard, K. G., Niyogi, D., Bonan, G., Lawrence, P., Baker, B., McNider, R., McAlpine, C., Etter, A., Gameda, S., Qian, B., Carleton, A., Beltran-Przekurat, A., Chase, T., Quintanar, A. I., Adegoke, J. O., Vezhapparambu, S., Conner, G., Asefi, S., Sertel, E., Legates, D. R., Wu, Y., Hale, R., Frauenfeld, O. W., Watts, A., Shepherd, M., Mitra, C., Anantharaj, V. G., Fall, S., Lund, R., Treviño, A., Blanken, P., Du, J., Chang, H.-I., Leeper, R., Nair, U. S., Dobler, S., Deo, R. & Syktus, J. (2010). Impacts of land use/land cover change on climate and future research priorities, *Bull. Amer. Meteor. Soc.*, **91**, 37–46, doi:10.1175/2009BAMS2769.1.

Mayer-Gürr, T. (2006). *Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE*, Ph.D. thesis, University of Bonn.

Mayer-Gürr, T., Zehentner, N., Klinger, B. & Kvas, A. (2014). ITSG-GRACE2014: A new GRACE gravity field release computed in Graz, GRACE Sci. Team Meet., Potsdam.

McKnight, T. L. & Hess, D. (2000). Climate zones and types: Climate zones and types, *Physical geography: A landscape appreciation. Upper Saddle River, NJ: Prentice Hall*, 223–6.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D. & Stockdale, T. (2009). Decadal prediction. can it be skillful?, *Bull. Amer. Meteor. Soc.*, **90**, 1467–1485.

Milly, P. C. D., Dunne, K. A. & Vecchia, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, **438**, 347–350, doi:10.1038/nature04312.

Müller, W. A., Baehr, J., Haak, H., Jungclaus, J. H., Kröger, J., Matei, D., Notz, D., Pohlmann, H., von Storch, J. S. & Marotzke, J. (2012). Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, *Geophys. Res. Lett.*, **39**, doi:10.1029/2012GL053326.

Müller, W. A., Pohlmann, H., Sienz, F. & Smith, D. (2014). Decadal climate predictions for the period 1901-2010 with a coupled climate model, *Geophysical Research Letters*, **41**, 2100–2107, doi:10.1002/2014GL059259, URL http://dx.doi.org/10.1002/2014GL059259, 2014GL059259.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, *Hydrol. Earth Syst. Sci.*, **18**, 3511–3538, doi:10.5194/hess-18-3511-2014, URL http://www.hydrol-earth-syst-sci.net/18/3511/2014/.

Niu, G.-Y. & Yang, Z.-L. (2006). Assessing a land surface model's improvements with grace estimates, *Geophysical Research Letters*, **33**, n/a–n/a, doi:10.1029/2005GL025555, URL http://dx.doi.org/10.1029/2005GL025555, l07401.

Pail, R., Bingham, R., Braitenberg, C., Dobslaw, H., Eicker, A., Güntner, A., Horwath, M., Ivins, E., Longuevergne, L., Panet, I. & Wouters, B. (2015). Science and user needs for observing global mass transport to understand global change and to benefit society, *Surveys in Geophysics*, **36**, 743–772, doi:10.1007/s10712-015-9348-9, URL http://dx.doi.org/10.1007/s10712-015-9348-9.

Palmer, M. A., Reidy Liermann, C. A., Nilsson, C., Flörke, M., Alcamo, J., Lake, P. S. & Bond, N. (2008). Climate change and the world's river basins: anticipating management options, *Frontiers in Ecology and the Environment*, **6**, 81–89, doi:10.1890/060148, URL http://dx.doi.org/10.1890/060148.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J. & Wood, E. F. (2012). Multisource estimation of long-term terrestrial water budget for major global river basins, *Journal of Climate*, **25**, 3191–3206, doi:10.1175/JCLI-D-11-00300.1, URL http://dx.doi.org/10.1175/JCLI-D-11-00300.1.

Paulson, A., Zhong, S. & Wahr, J. (2007). Inference of mantle viscosity from GRACE and relative sea level data, *Geophysical Journal International*, **171**, 497–508, doi:10.1111/j.1365-246X.2007.03556.x, URL http://dx.doi.org/10.1111/j.1365-246X.2007.03556.x.

Petit, G. & Luzum, B. (2010). *IERS Conventions (2010)*, *IERS Technical Note ; 36*, Bundesamt für Kartographie und GeodÃ¤sie, Frankfurt am Main.

Phillips, T., Nerem, R. S., Fox-Kemper, B., Famiglietti, J. S. & Rajagopalan, B. (2012). The influence of ENSO on global terrestrial water storage using GRACE, *Geophysical Research Letters*, **39**, n/a–n/a, doi:10.1029/2012GL052495, URL http://dx.doi.org/10.1029/2012GL052495, l16705.

Pierce, D., Barnett, T., Tokmakian, R., Semtner, A., Maltrud, M., Lysne, J. & Craig, A. (2004). The ACPI Project, Element 1: Initializing a Coupled Climate Model from Observed Conditions, *Clim. Chang.*, **62**, 13–28, doi:10.1023/B:CLIM.0000013676.42672.23.

Pohlmann, H., Müller, W. A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F. S. E., Kadow, C., Illing, S. & Marotzke, J. (2013). Improved forecast skill in the tropics

in the new MiKlip decadal climate predictions, *Geophys. Res. Lett.*, **40**, 5798–5802, doi:10.1002/2013GL058051, URL `http://dx.doi.org/10.1002/2013GL058051`.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K. G., Wetzel, P. & Jungclaus, J. (2007). Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, *Clim. Dyn.*, **29**, 565–574, doi:10.1007/s00382-007-0247-8.

Räisänen, J. & Ylhäisi, J. S. (2011). How much should climate model output be smoothed in space?, *Journal of Climate*, **24**, 867–880, doi:10.1175/2010JCLI3872.1.

Ramillien, G., Frappart, F., Güntner, A., Ngo-Duc, T., Cazenave, A. & Laval, K. (2006). Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, *Water Resources Research*, **42**, n/a–n/a, doi:10.1029/2005WR004331, URL `http://dx.doi.org/10.1029/2005WR004331`, w10403.

Reager, J. T., Thomas, B. F. & Famiglietti, J. S. (2014). River basin flood potential inferred using GRACE gravity observations at several months lead time, *Nat. Geosci.*, **7**, 588–592, doi:10.1038/ngeo2203.

Reick, C. H., Raddatz, T., Brovkin, V. & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM, *JAMES*, **5**, 459–482, doi:10.1002/jame.20022, URL `http://dx.doi.org/10.1002/jame.20022`.

Rietbroek, R., Fritsche, M., Brunnabend, S.-E., Daras, I., Kusche, J., Schröter, J., Flechtner, F. & Dietrich, R. (2012). Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data, *Journal of Geodynamics*, **59-60**, 64 – 71, doi:http://dx.doi.org/10.1016/j.jog.2011.02.003, URL `http://www.sciencedirect.com/science/article/pii/S0264370711000305`, mass Transport and Mass Distribution in the System Earth.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K. & Cosgrove, B. (2004). The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, **85**, 381–394.

Rodell, M., Velicogna, I. & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in india, *Nature*, **460**, 999–1002, doi:10.1038/nature08238.

Rowlands, D. D., Luthcke, S. B., Klosko, S. M., Lemoine, F. G. R., Chinn, D. S., McCarthy, J. J., Cox, C. M. & Anderson, O. B. (2005). Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements, *Geophysical Research Letters*, **32**, n/a–n/a, doi:10.1029/2004GL021908, URL `http://dx.doi.org/10.1029/2004GL021908`, l04310.

Rowlands, D. D., Luthcke, S. B., McCarthy, J. J., Klosko, S. M., Chinn, D. S., Lemoine, F. G., Boy, J. & Sabaka, T. J. (2010). Global mass flux solutions from GRACE: A comparison of parameter estimation strategies – Mass concentrations versus Stokes coefficients, *Journal of Geophysical Research (Solid Earth)*, **115**, 1403–+, doi:10.1029/2009JB006546.

Sasgen, I., Dobslaw, H., Martinec, Z. & Thomas, M. (2010). Satellite gravimetry observation of Antarctic snow accumulation related to ENSO, *Earth Planet Sc. Lett.*, **299**, 352–358, doi:10.1016/j.epsl.2010.09.015.

Savcenko, R. & Bosch, W. (2008). *EOT08a-empirical ocean tide model from multi-mission satellite altimetry*, *Scientific technical report*, Deutsches Geodätisches Forschungsinstitut (DGFI), München.

Savcenko, R. & Bosch, W. (2012). *EOT11a - Empirical Ocean Tide Model from Multi-Mission Satellite Altimetry*, *DGFI Report No. 89*, Deutsches GeodÃ¤tisches Forschungsinstitut (DGFI), München.

Schewe, J., Heinke, J., Gerten, D., Hadeland, I., Arnell, N. w., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X. C., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q. H., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Pointek, F., Warszawski, L. & Kabat, P. (2014). Multimodel assessment of water scarcity under climate change, *PNAS*, **111**, 3245–3250., doi:10.1073/pnas.1222460110.

Schmidt, R., Flechtner, F., König, R., Meyer, U., Neumayer, K. H., Reigber, C., Rothacher, M., Petrovic, S., Zhu, S. Y. & Güntner, A. (2007). *GRACE Time-Variable Gravity Accuracy Assessment*, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 237–243, doi:10.1007/978-3-540-49350-1$_3$6, URL `http://dx.doi.org/10.1007/978-3-540-49350-1_36`.

Schmidt, R., Flechtner, F., Meyer, U., Neumayer, K.-H., Dahle, C., König, R. & Kusche, J. (2008). Hydrological signals observed by the GRACE satellites, *Surveys in Geophysics*, **29**, 319–334, doi:10.1007/s10712-008-9033-3, URL `http://dx.doi.org/10.1007/s10712-008-9033-3`.

Schmidt, T., Schoon, L., Dobslaw, H., Matthes, K., Thomas, M. & Wickert, J. (2015). Utls temperature validation of mpi-esm decadal hindcast experiments with gps radio occultations, *Meteorologische Zeitschrift*, –, doi:10.1127/metz/2015/0601, URL `http://dx.doi.org/10.1127/metz/2015/0601`.

Seneviratne, S., Viterbo, P., Lüthi, D. & Schär, C. (2004). Inferring changes in terrestrial water storage using ERA-40 reanalysis data : the Mississippi River basin, *J. Climate*, **17**, 2039–2057, doi:10.1175/1520-0442(2004)017<2039:ICITWS>2.0.CO;2.

Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Oleson, K. W. & Verseghy, D. (2006). Soil Moisture Memory in AGCM Simulations: Analysis of Global Land-Atmosphere Coupling Experiment (GLACE) Data, *Journal of Hydrometeorology*, **7**, 1090–1112, doi:10.1175/JHM533.1, URL `http://dx.doi.org/10.1175/JHM533.1`.

Seneviratne, S. I. & Stöckli, R. (2007). *The role of land-atmosphere interactions for climate variability in Europe*, *Advances in Global Change Research*, vol. 33, Springer Netherlands, 179-193 p., doi:10.1007/978-1-4020-6766-2$_1$2, URL `http://dx.doi.org/10.1007/978-1-4020-6766-2_12`.

Seneviratne, S. I., Wilhelm, M., Stanelle, T., van den Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M. E., Meier, A., Brovkin, V., Claussen, M., Ducharne, A., Dufresne,

J.-L., Findell, K. L., Ghattas, J., Lawrence, D. M., Malyshev, S., Rummukainen, M. & Smith, B. (2013). Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment, *Geophys. Res. Lett.*, **40**, 5212–5217, doi:10.1002/grl.50956, URL `http://dx.doi.org/10.1002/grl.50956`.

Seo, K.-W., Wilson, C. R., Famiglietti, J. S., Chen, J. L. & Rodell, M. (2006). Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE), *Water Resources Research*, **42**, n/a–n/a, doi:10.1029/2005WR004255, URL `http://dx.doi.org/10.1029/2005WR004255`, w05417.

Smith, D., Eade, R. & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction, *Clim. Dyn.*, **41**, 3325–3338, doi:10.1007/s00382-013-1683-2, URL `http://dx.doi.org/10.1007/s00382-013-1683-2`.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R. & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model, *Science*, **317**, 796–799, doi:10.1126/science.1139540, URL `http://www.sciencemag.org/content/317/5839/796.abstract`.

Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H. & Scaife, A. A. (2010). Skilful multi-year predictions of Atlantic hurricane frequency, *Nat. Geosci.*, **3**, 846–849, doi:10.1038/ngeo1004, URL `http://dx.doi.org/10.1038/ngeo1004`.

Smith, D. M., Scaife, A. A., Eade, R. & Knight, J. R. (2016). Seasonal to decadal prediction of the winter north atlantic oscillation: emerging capability and future prospects, *Quarterly Journal of the Royal Meteorological Society*, **142**, 611–617, doi:10.1002/qj.2479, URL `http://dx.doi.org/10.1002/qj.2479`.

Spangehl, T., Schröder, M., Stolzenberger, S., Glowienka-Hense, R., Mazurkiewicz, A. & Hense, A. (2015). Evaluation of the MiKlip decadal prediction system using satellite based cloud products, *Meteorol. Z.*, URL `http://dx.doi.org/10.1127/metz/2015/0602`, [Online; posted 04-April-2015].

Stacke, T. & Hagemann, S. (2012). Development and evaluation of a global dynamical wetlands extent scheme, *Hydrol. Earth Syst. Sci.*, **16**, 2915–2933, doi:10.5194/hess-16-2915-2012.

Standish, E. (1998). *JPL planetary and lunar ephemerides DE405/LE405, IOM 312.F-98-048*, *Scientific technical report*, JPL, Los Angeles, USA.

Steffen, H., Denker, H. & M
uller, J. (2008). Glacial isostatic adjustment in Fennoscandia from GRACE data and comparison with geodynamical models, *Journal of Geodynamics*, **46**, 155–164, doi:http://dx.doi.org/10.1016/j.jog.2008.03.002.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T. & Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6, *JAMES*, **5**, 146–172, doi:10.1002/jame.20015.

Swenson, S., Chambers, D. & Wahr, J. (2008). Estimating geocenter variations from a combination of GRACE and ocean model output, *J. Geophys. Res., B*, **113**, 8410–+, doi:10.1029/2007JB005338.

Swenson, S. & Wahr, J. (2002). Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *J. Geophys. Res., B*, **107**, ETG 3–1–ETG 3–13, doi:10.1029/2001JB000576.

Swenson, S. & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, **33**, doi:10.1029/2005GL025285, l08402.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J. & Wilson, C. R. (2008a). Analysis of terrestrial water storage changes from GRACE and GLDAS, *Water Resour. Res.*, **44**, doi:10.1029/2006WR005779, W02433.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J. & Wilson, C. R. (2008b). Analysis of terrestrial water storage changes from GRACE and GLDAS, *Water Resour. Res.*, **44**, doi:10.1029/2006WR005779, URL `http://dx.doi.org/10.1029/2006WR005779`.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. & Watkins, M. M. (2004a). GRACE measurements of mass variability in the Earth system, *Science*, **305**, 503–505, doi:10.1126/science.1099192, URL `http://science.sciencemag.org/content/305/5683/503`.

Tapley, B. D., Bettadpur, S., Watkins, M. & Reigber, C. (2004b). The gravity recovery and climate experiment: Mission overview and early results, *Geophys. Res. Lett.*, **31**, doi:10.1029/2004GL019920, L09607.

Tapley, B. D., Flechter, F., Watkins, M. & Bettadpur, S. (2014). GRACE Mission: Status and Prospects, GRACE Science Team Meeting (GSTM), Potsdam, Germany.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design, *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thomas, M., Sündermann, J. & Maier-Reimer, E. (2001). Consideration of ocean tides in an ogcm and impacts on subseasonal to decadal polar motion excitation, *Geophysical Research Letters*, **28**, 2457–2460, doi:10.1029/2000GL012234, URL `http://dx.doi.org/10.1029/2000GL012234`.

Tiwari, V. M., Wahr, J. & Swenson, S. (2009). Dwindling groundwater resources in northern india, from satellite gravity observations, *Geophysical Research Letters*, **36**, n/a–n/a, doi:10.1029/2009GL039401, URL `http://dx.doi.org/10.1029/2009GL039401`, l18401.

Tregoning, P., Watson, C., Ramillien, G., McQueen, H. & Zhang, J. (2009). Detecting hydrologic deformation using GRACE and GPS, *Geophysical Research Letters*, **36**, n/a–n/a, doi:10.1029/2009GL038718, URL `http://dx.doi.org/10.1029/2009GL038718`, l15401.

Troch, P., Durcik, M., Seneviratne, S., Hirschi, M., Teuling, A., Hurkmans, R. & Hasan, S. (2007). New data sets to estimate terrestrial water storage change, *Eos, Trans. Amer.*, **88**, 469–470, doi:10.1029/2007EO450001.

Valcke, S. (2013). The OASIS3 coupler: a European climate modelling community software, *Geoscientific Model Development*, **6**, 373–388, doi:10.5194/gmd-6-373-2013, URL `http://www.geosci-model-dev.net/6/373/2013/`.

van Dam, T., Wahr, J. & Lavallée, D. (2007). A comparison of annual vertical crustal displacements from GPS and Gravity Recovery and Climate Experiment (GRACE) over Europe, *Journal of Geophysical Research: Solid Earth*, **112**, n/a–n/a, doi:10.1029/2006JB004335, URL `http://dx.doi.org/10.1029/2006JB004335`, b03404.

Velicogna, I. & Wahr, J. (2006). Measurements of time-variable gravity show mass loss in Antarctica, *Science*, **311**, 1754–1756, doi:10.1126/science.1123785, URL `http://science.sciencemag.org/content/311/5768/1754`.

von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernández-Deckers, D., Maier-Reimer, E., Marotzke, J. & Stammer, D. (2012). An estimate of the lorenz energy cycle for the world ocean based on the STORM/NCEP simulation, *Journal of Physical Oceanography*, **42**, 2185–2205, doi:10.1175/JPO-D-12-079.1, URL `http://dx.doi.org/10.1175/JPO-D-12-079.1`.

Vörösmarty, C., Fekete, B., Meybeck, M. & Lammers, R. (2000b). A simulated topological network representing the global system of rivers at 30-minute spatial resolution (STN-30), *Global Biogeochem. Cy.*, **14**, 599–621.

Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M. & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, *Water Resour. Res.*, **49**, 904–914, doi:10.1002/wrcr.20078.

Wahr, J. (2009). *Time-variable gravity from satellites*, vol. 3, Elsevier, 213-237 p.

Wahr, J., Molenaar, M. & Bryan, F. (1998). Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res., B*, **103**, 30,205–30,229, doi:10.1029/98JB02844.

Wahr, J., Swenson, S. & Velicogna, I. (2006). Accuracy of GRACE mass estimates, *J. Geophys. Res.*, **33**, doi:10.1029/2005GL025305, L06401.

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C. & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, *Journal of Geophysical Research: Solid Earth*, **120**, 2648–2671, doi:10.1002/2014JB011547, URL `http://dx.doi.org/10.1002/2014JB011547`, 2014JB011547.

Watkins, M. M., Yuan, D., Kuang, D., Bertiger, W., Kim, M. & Kruizinga, G. L. (2005). GRACE Harmonic and Mascon Solutions at JPL, American Geophysical Union (AGU), San Francisco, Calif., abstract G22A-04.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, **50**, 7505–7514, doi:10.1002/2014WR015638, URL `http://dx.doi.org/10.1002/2014WR015638`.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., & Best, M. (2011). Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, *J. Hydrometeor*, **12**, 823–848, doi:0.1175/2011JHM1369.1.

Werth, S. (2010). *Calibration of the global hydrological model WGHM with water mass variations from GRACE gravity data*, Ph.D. thesis, Univ. Potsdam.

Werth, S., Güntner, A., Schmidt, R. & Kusche, J. (2009). Evaluation of GRACE filter tools from a hydrological perspective, *Geophys. J. Int.*, **179**, 1499–1515, doi:10.1111/j.1365-246X.2009.04355.x.

Widén-Nilsson, E., Halldin, S. & yu Xu, C. (2007). Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, *Journal of Hydrology*, **340**, 105 – 118, doi:http://dx.doi.org/10.1016/j.jhydrol.2007.04.002, URL `http://www.sciencedirect.com/science/article/pii/S0022169407002120`.

Wiese, D. N., Nerem, R. S. & Lemoine, F. G. (2012). Design considerations for a dedicated gravity recovery satellite mission consisting of two pairs of satellites, *Journal of Geodesy*, **86**, 81–98, doi:10.1007/s00190-011-0493-8, URL `http://dx.doi.org/10.1007/s00190-011-0493-8`.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, International geophysics series, Academic Press, Oxford, Waltham (Mass.), URL `http://opac.inria.fr/record=b1135201`.

Wolff, M. (1969). Direct measurements of the Earth's gravitational potential using a satellite pair, *Journal of Geophysical Research*, **74**, 5295–5300, doi:10.1029/JB074i022p05295, URL `http://dx.doi.org/10.1029/JB074i022p05295`.

Wouters, B., Bonin, J. A., Chambers, D. P., Riva, R. E. M., Sasgen, I. & Wahr, J. (2014). GRACE, time-varying gravity, Earth system dynamics and climate change, *Reports on Progress in Physics*, **77**, 116,801, URL `http://stacks.iop.org/0034-4885/77/i=11/a=116801`.

Xie, P. & Arkin, P. A. (1997). Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Zaitchik, B. F., Rodell, M. & Reichle, R. H. (2008). Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin, *Journal of Hydrometeorology*, **9**, 535–548, doi:10.1175/2007JHM951.1, URL `http://dx.doi.org/10.1175/2007JHM951.1`.

Zeng, N., Yoon, J.-H., Mariotti, A. & Swenson, S. (2008). Variability of basin-scale terrestrial water storage from a PER water budget method: The Amazon and the Mississippi, *J. Climate*, **21**, 248–265, doi:10.1175/2007JCLI1639.1.

Zenner, L., Bergmann-Wolf, I., Dobslaw, H., Gruber, T., Güntner, A., Wattenbach, M., Esselborn, S. & Dill, R. (2014). Comparison of daily GRACE gravity field and numerical water storage models for de-aliasing of satellite gravimetry observations, *Surveys in Geophysics*, **35**, 1251–1266, doi:10.1007/s10712-014-9295-x, URL `http://dx.doi.org/10.1007/s10712-014-9295-x`.

Zhang, L., Dobslaw, H., Dahle, C., Sasgen, I. & Thomas, M. (2015). Validation of MPI-ESM decadal hindcast experiments with terrestrial water storage variations as observed by the GRACE satellite mission, *Meteorologische Zeitschrift*, –, doi:10.1127/metz/2015/0596, URL `http://dx.doi.org/10.1127/metz/2015/0596`.

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R. & Thomas, M. (2017). Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, *Hydrology and Earth System Sciences*, **21**, 821–837, doi:10.5194/hess-21-821-2017, URL `http://www.hydrol-earth-syst-sci.net/21/821/2017/`.

Zhang, L., Dobslaw, H. & Thomas, M. (2016). Globally gridded terrestrial water storage variations from GRACE satellite gravimetry for hydrometeorological applications, *Geophysical Journal International*, **206**, 368–378, doi:10.1093/gji/ggw153, URL `http://gji.oxfordjournals.org/content/206/1/368.abstract`.

Zhang, S. (2011). Impact of observation-optimized model parameters on decadal predictions: Simulation with a simple pycnocline prediction model, *Geophysical Research Letters*, **38**, n/a–n/a, doi:10.1029/2010GL046133, URL `http://dx.doi.org/10.1029/2010GL046133`, l02702.

Zuber, M. T., Smith, D. E., Watkins, M. M., Asmar, S. W., Konopliv, A. S., Lemoine, F. G., Melosh, H. J., Neumann, G. A., Phillips, R. J., Solomon, S. C., Wieczorek, M. A., Williams, J. G., Goossens, S. J., Kruizinga, G., Mazarico, E., Park, R. S. & Yuan, D.-N. (2013). Gravity field of the moon from the Gravity Recovery and Interior Laboratory (GRAIL) mission, *Science*, **339**, 668–671, doi:10.1126/science.1231507, URL `http://science.sciencemag.org/content/339/6120/668`.

Acknowledgements

I want to express my great thanks to Maik Thomas and Henryk Dobslaw, for their trust, patience and instructions during the years of my PhD study. Maik is always supportive whenever I am in need of help and affects me with his research attitude, hard working and great knowledge. Henryk is a great supervisor. He cares about my work, helps me with my problems, and also leaves me freedom to investigate into the problems by myself. I wish to thank him for his supervision, encouragement and strict requirement.

I would like also to thank all my former and current colleagues within our section for being kind and helpful to me. Special thanks would go to Inga Bergmann-Wolf and Robert Dill, for a lot of technical and scientific supports. Inga helps me a lot during my study. Thanks for her patience when I come to her with some problems. Thanks to Ronny Wolf for the technique assistance. And thanks to my office mate, Lea, for being such a nice friend to me and helpful in both my research and personal life. Besides, I would also thank some colleagues outside our section, Andreas Güntner, Christoph Dahle, and Tobias Stacke for their scientific support and discussions, which are of great importance to me. And I also want to thank Frauke Stobbe from Welcome Center Potsdam, for all her kind help that solves a lot of difficult troubles for us and makes our life much easier in Germany.

Thanks to my husband Minghui, my daughter Wuyou and my parents for providing me a family with love.

This work has been supported by the German Federal Ministry of Education and Research (BMBF) within the FONA research program under the grants 03F0654A and 01LP1151A.